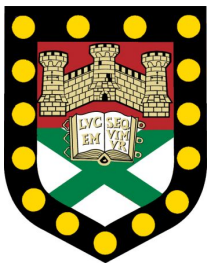
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Open Research Exeter

Work, entropy and uncertainties in thermodynamics beyond the classical and weak coupling regime



Number of volumes: 1

Submitted by Harry Miller, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics, April 2019.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Abstract

Thermodynamics typically concerns the physical behaviour of macroscopic systems comprised of many particles. However, recent theoretical progress has extended the theory to both the classical-stochastic and quantum regimes, where systems are comprised of just a small number of particles. In this thesis I investigate a range of situations in which new thermodynamic phenomena emerge due to the reduced size of the systems involved. One central assumption in macroscopic thermodynamics is the weak coupling approximation, which posits that the equilibrium properties of a system are not influenced by the interactions with its surrounding environment. However, for microscopic systems this assumption can break down, and I derive new fluctuation relations that provide a refined form of the second law of thermodynamics in this strong-coupling regime, taking into account corrections stemming from these interactions. In this work I provide a characterisation of stochastic heat and entropy production for small scale classical systems that are defined regardless of the strength of interaction. I then show that these definitions lead to a consistent thermodynamic framework valid beyond the usual weak-coupling regime. The thesis also concerns the effect of interactions on the equilibrium properties of strongly-coupled quantum systems, and I investigate how these interactions can influence the resulting temperature fluctuations in this regime. Using tools from quantum estimation theory, I derive an uncertainty relation between energy and temperature valid at all coupling strengths and system sizes. The relation reveals how quantum energy coherences contribute to statistical fluctuations in the estimated temperature of small-scale systems. Finally, I investigate how quantum fluctuations influence the statistics of work extracted from slowly-driven quantum systems. I prove that, unlike in classical systems, the work dissipated by a general quantum system is no longer proportional to its statistical fluctuations. This result reveals new subtleties involved in designing optimal quantum thermodynamic processes.

Author's declaration

The contents of this thesis contains work done in collaboration with other authors from the following papers:

- (*Chapter 2*): H. J. D. Miller & J. Anders, " Entropy production and time asymmetry in the presence of strong interactions", Physical Review E 95, 062123 (2017)

In this published work I derived all mathematical results of the paper and contributed equally to the final discussion and interpretation of the results.

- (*Chapter 3*): H. J. D. Miller & J. Anders, "Energy-temperature uncertainty relation in quantum thermodynamics", Nature Communications 9, 2203 (2018)

In this published work I derived all main results of the paper, apart from numerical calculations performed by my co-author. I then contributed equally to the final discussion and interpretation of the results.

- (*Chapter 4*): H. J. D. Miller, M. Scandi, J. Anders & M. Perarnau-Llobet, "Work fluctuations in slow processes: quantum signatures and optimal control" (unpublished, arXiv:1905.07328) (2019).

In this unpublished work I contributed equally to the derivations of the main results, as well as the discussion and writing of the final draft.

Throughout the thesis all theorems, lemmas and proofs represent my own individual contributions to these works unless otherwise stated.

Table of Contents

List of Figures

Figure **Page**

Chapter 1

Introduction

Despite the many paradigm shifts in physics over the last 150 years, thermodynamics has proved to be one of the most robust physical theories in modern science. While it initially grew out of a necessity to design efficient heat engines in the early 18th century, pioneering work by Boltzmann and Gibbs developed thermodynamics into a fundamental theory which has become crucial to our understanding of irreversibility in nature and the flow of time. As Einstein famously remarked, "Thermodynamics is the only physical theory which I am convinced will never be overthrown, within the framework of applicability of its basic concepts." [1].

One of those basic concepts is the notion of weak coupling. If we consider a macroscopic three-dimensional body interacting with its environment, then simple intuition suggests that the energy contained in this interaction should be sufficiently negligible. This is because interactions will typically only occur along the body's surface boundary, involving only a small fraction of the atoms making up the system. Compared with the energy contained in the bulk material this is negligible, and the energetics of the system and environment can be partitioned separately. In the standard formulation of thermodynamics this weak-coupling assumption is crucial for a consistent description of heat flow, as any energy lost by the system during a process can simply be equated to energy dissipated into the environment in the form of heat. In addition to energy, this assumption is directly related to fundamental properties of the thermodynamic entropy and heat capacity, as both quantities become extensive with respect to the volume of the system [2].

While traditionally thermodynamics has always been concerned with describing macroscopic systems consisting of a large number of particles, advancements in nano- and meso-scopic physics demonstrate that small systems, such as biopolymers and colloidal particles, may exhibit non-trivial thermal behaviour even at sub-micron scales. For example, individual protein molecules can function as tiny motors that convert chemical energy into mechanical work [3], and single strands of RNA that are manipulated using optical tweezers exhibit dissipation and lag much like any macroscopic system governed by the laws of thermodynamics [4]. One may even conceive of quantum thermal machines constructed from individual few-level systems [5], and the analysis regarding the transport of heat and energy at the quantum scale suggests that thermodynamics can be consistently extended beyond the classical, macroscopic setting [6]. At these length scales both stochastic and quantum fluctuations begin to dominate, and current research is now devoted to investigating the potential modifications to thermodynamics that occur at or below the nanoscale. In this thesis I explore various avenues in which the traditional theory of thermodynamics breaks down in the microscopic regime, while presenting new theoretical approaches that extend the theory beyond its traditional assumptions.

One major modification is the break-down of the weak coupling approximation. It becomes clear that the heuristic argument used to justify this assumption fails to apply to systems with smaller volumes. If the surface area of the system becomes comparable to its volume then the energetic contributions to the interaction with the environment also become comparable to its own internal energy. When this occurs the thermodynamic properties of the system become dependent on the nature of the interaction, both in and out of equilibrium [7]. One clear example of this behaviour occurs in the collapse transition of a polymer immersed in a solvent, which can become significantly dependent on the structure of the solvent itself rather than just the surrounding temperature [8]. Another example concerns a quantum harmonic oscillator in a black-body radiation field, whereby the free energy of the oscillator shifts by a temperature and coupling-dependent term due to a non-vanishing dipole interaction with the field [9].

In such situations special care needs to taken when defining notions of heat, work and entropy, as the thermal properties of these small systems cannot generally be determined simply by their respective internal Hamiltonians. In Chapter 2 it will be shown that this issue stems from the fact that the equilibrium configuration of the system can deviate from the familiar Gibbs distribution when interactions are non-negligible. This invalidates the usual formulas used to compute the associated thermodynamic potentials, such as the free energy, as these are derived from the assumption that the equilibrium state of the

system is of Gibbs form. This of course does not mean that thermodynamics breaks down. Instead, one can construct alternative thermodynamic potentials via an effective operator or function known as the *Hamiltonian of mean force* (HMF). Much like standard methods in statistical mechanics, this mathematical object can be used to compute an effective partition function for the system, which provides the basis for defining the system's free energy and all resulting equilibrium properties. This effective description of the system in terms of the modified partition function dates back to the early work of Kirkwood in his analysis of the statistical properties of fluid mixtures [10]. I will discuss the role played by the HMF in extending thermodynamics beyond the weak-coupling regime in classical thermodynamics, both in and away from equilibrium. We will see that the HMF provides a consistent generalisation of the laws of thermodynamics and fluctuation relations, and can be used to define entropy production and quantify time-asymmetry for out-of-equilibrium processes. Chapter 2 will relate to my published work on this topic [11], where I formulate this extension of non-equilibrium thermodynamics to the strong-coupling regime.

In Chapter 3 I focus on quantum mechanical systems in the strong coupling regime. While in classical thermodynamics all physical quantities are determined from a microscopic description based on phase space, in the quantum regime thermodynamics can be formulated from an underlying Hilbert space structure instead. Due to the non-commutativity between certain observables in open quantum systems, additional subtleties arise when trying to define thermodynamic variables and their corresponding statistical fluctuations [12]. This problem is most apparent in the strong coupling regime, as the presence of a finite interaction between system and reservoir can result in a system state that is non-diagonal in its own energy basis [13]. In this chapter I explore the consequences of these strong-coupling coherences in the context of temperature estimation. By using the quantum generalisation of the Hamiltonian of mean force, I develop a new framework for quantifying non-classical energy fluctuations in strongly-coupled quantum systems, utilising a measure of quantum uncertainty known as the Wigner-Yanase-Dyson skew information [14]. In particular, I demonstrate, via a new form of equilibrium fluctuation-dissipation relation, that the heat capacity of a quantum system relates directly to the amount of quantum energy fluctuations contained in the state. Using this framework I provide a fully-quantum generalisation of the energy-temperature uncertainty relation. Classically, this uncertainty relation represents a trade-off between statistical fluctuations in the temperature and energy of weakly-coupled systems [2]. Using techniques from estimation theory, I find additional correction terms to this uncer-

tainty relation in the quantum strong-coupling setting, and show that energy coherences contribute to the fluctuations in temperature of quantum systems below the nanoscale. These results are based on my published work [15]. In the final part of this chapter I consider the implications of these results for quantum thermometry, whereby one can use a small quantum system as a probe in order to ascertain the temperature of a larger many-body system in thermal equilibrium. In particular, I derive tighter constraints on the performance of a quantum thermometer connected to a bosonic substance by taking into account the presence of strong interactions between probe and sample. Overall, the results of Chapter 3 demonstrate that the presence of non-negligible interactions in quantum thermodynamics leads to new physical behaviour that must be accounted for in any accurate thermodynamic description of a microscopic system.

In the final part of thesis, Chapter 4 concerns the role of quantum fluctuations on the statistics of work extraction from quantum systems. In the quantum regime work is not an observable, and is instead defined from measurements of transitions between energy eigenstates of the driven system [16]. In this framework one can construct a quantum version of stochastic thermodynamics, with the standard fluctuation relations all fulfilled along with the first and second laws of thermodynamics [17]. However, despite the wealth of classical results that have been shown to hold in the quantum setting, the so-called work fluctuation-dissipation relation [18] has not been previously derived for quantum systems. This relation shows that in the slow-driving limit, the work dissipated by any driven system close to equilibrium should be proportional to the resulting stochastic fluctuations in work. In this Chapter I treat the same situation quantum-mechanically, and show that this work fluctuation-dissipation relation actually breaks down. Instead, I derive a new relation between work and its fluctuations that demonstrates how quantum fluctuations in power modify the standard classical result. In particular, it is proven that rather than being proportional, work fluctuations actually exceed the amount of dissipation in general quantum systems driven close to equilibrium. I further demonstrate that this uniquely non-classical behaviour can be witnessed experimentally. In the final part of the chapter I connect these results to the field of thermodynamic geometry [19]. In classical thermodynamics, one may represent out-of-equilibrium processes in terms of Riemannian geometry, and this can be used to find optimal protocols that simultaneously minimise both work dissipation and fluctuations. However, I show that in the quantum regime a different geometric structure emerges in which the path of minimal dissipation is not equivalent to the path of minimal work fluctuations. Instead, one must resort to a compromise between the two quantities. This important observation stems directly from

the presence of non-classical fluctuations in power.

Chapter 2

Stochastic thermodynamics beyond weak-coupling

Thermodynamics is a theory concerning the physical behaviour of macroscopic objects consisting of many interacting particles. Its laws describe the partitioning of energy exchanges into heat and useful work, with entropy playing the pivotal role as a quantity determining whether or not a thermodynamic process is possible. In essence, thermodynamics itself is a purely phenomenological theory, but its laws can be proven microscopically using Hamiltonian mechanics [2]. However, due to the large numbers of particles involved in such processes, full information about a macroscopic body can never be obtained. Thus whilst the motion of each individual particle is purely deterministic according to Hamilton's equations of motion, predictions about the exchanges of energy and entropy will be subject to stochastic fluctuations due to an observer's lack of information about the state of the evolving system. In an appropriate thermodynamic limit, these fluctuations effectively disappear, and one can establish the phenomenological laws of thermodynamics at the macroscopic scale.

However, as the system size is scaled down, microscopic fluctuations in energy and entropy become appreciable outside of the thermodynamic limit. In this regime, one can still predict the ensemble behaviour of a finite system, though changes in average energy and entropy go hand in hand with non-negligible statistical uncertainties. Understanding the physical behaviour of these fluctuations, along with their thermodynamic implications for the control of nanoscale devices, is the central goal of *stochastic thermodynamics*. The aim here is to provide a microscopic description of entropy production at the level of the individual trajectories traced out by the system as it is driven away from equilibrium

[20–24]. Entropy production provides us with a quantitative description of change and irreversibility in nature, and its average increase places restrictions on allowed state transformations in accordance with the second law of thermodynamics [2, 25]. However, in this stochastic regime one can obtain more refined statements about the nature of entropy production using the so-called *fluctuation theorems* [18, 21, 26–29]. These relations go beyond the standard predictions of thermodynamics and provide universal insight into the breaking of time-reversal symmetry in a wide variety of physical systems [4, 30–33]. By analysing the statistical behaviour of entropy production, one can gain vital insight into the connection between thermodynamics and information theory [24].

Standard analysis of entropy production in open systems, both quantum and classical, centres on an assumption that the system *weakly* interacts with a thermal reservoir [24, 34–36]. The benefit of this assumption is that it provides an unambiguous notion of stochastic heat, since neglecting energetic contributions from the interaction provides a clear division between the energy of the system and the reservoir. While the weak coupling assumption can be physically justified in macroscopic systems, the thermodynamic behaviour of small-scale systems may be strongly influenced by a non-negligible interaction with their environment [7]. Thus it is of paramount importance to explore extended notions of entropy production within the strong coupling regime, which will be the subject of this chapter.

The extension of thermodynamics to the strong coupling regime has been the subject of recent debate in the context of both classical [7, 12, 23, 37, 38] and quantum systems [39–45]. The central question revolves around the identification of thermodynamic potentials for the system at both the stochastic and ensemble level. An elegant solution to this problem, originally dating back to Kirkwood in 1935 [10], is to replace the isolated Hamiltonian of the system with an effective Hamiltonian that takes into account the non-negligible interaction and temperature of the environment. This allows one to define an *effective* internal energy, free energy and entropy for the system at equilibrium [12]

While this effective Hamiltonian approach adequately describes equilibrium thermodynamics in the strong-coupling regime, in this chapter I will adapt the framework in order to describe entropy production for more general stochastic processes operating away from equilibrium. It will be shown that the equilibrium properties of the thermal reservoir can be still be well-defined, even when the system is arbitrarily far from equilibrium and correlated with the reservoir. For out-of-equilibrium processes I will show that the average entropy production can be related to the Kullback-Leibler divergence, thus extending the link between thermodynamics and information theory beyond weak-coupling. In addition,

it will be proven that the stochastic entropy production satisfies a generalised fluctuation relation and can be used to quantify time-asymmetry of correlated non-equilibrium processes. These results naturally extend the familiar properties of entropy production in weakly-coupled systems to the strong coupling regime. The structure of the chapter is as follows. I will first introduce the framework used to describe classical thermodynamics in the weak-coupling regime, both in and away from equilibrium, focusing on the well-known fluctuation relations for entropy production. I will then present the results of my publication [11], which provides a generalisation of these concepts to the strong-coupling regime. Finally, I will connect these findings to other results in the literature concerning entropy production away from equilibrium.

2.1 Preliminaries: Statistical mechanics

Before delving into underlying framework of stochastic thermodynamics, we will first establish the underlying structure of statistical mechanics and its connection to information theory. Let us suppose that we have an isolated composite system, consisting of N particles. Let $z_t = [\vec{X}(t), \vec{Q}(t)]$ represent the collective phase space of all particles at some time t. Here the vector $\vec{X}(t) = (X_1(t), X_2(t), ...X_N(t))$ describes the positions of each particle, while $\vec{Q}(t) = (Q_1(t), Q_2(t), ...Q_N(t))$ the set of momenta. We assume that our composite system evolves according to a Hamiltonian $H(z_t; \lambda_t)$, where λ_t is some time-dependent control parameter. The positions and momenta of each particle are determined by Hamilton's equations of motion:

$$\begin{aligned} \dot{X}_i(t) &= \frac{\partial H}{\partial Q_i}, \\ \dot{Q}_i(t) &= -\frac{\partial H}{\partial X_i}, \qquad i = (1, 2, .., N). \end{aligned} \tag{2.1}$$

At any given time the statistical state of the system is described by a probability density function over the outcome space $\{z_t\}$, and we characterise this set $\mathcal{P}$ by the set of differentiable functions of the form

$$\mathcal{P} := \left\{ \varrho(z_t; t) \mid \varrho(z_t; t) \geq 0, \quad \int dz_t \ \varrho(z_t; t) = 1 \right\}, \tag{2.2}$$

where all integrals are taken over the full phase space $\{z_t\}$. The ensemble average of some function $f(z_t)$ is then calculated according to

$$\langle f \rangle := \int dz_t \ \varrho(z_t; t) f(z_t). \tag{2.3}$$

By assumption the composite system is isolated, the probability density $\varrho(z_t;t) \in \mathcal{P}$ is conserved along any phase space trajectory and hence obeys Liouville's equation [2]:

$$\frac{\partial \varrho}{\partial t} = \mathcal{L}[\varrho], \tag{2.4}$$

with $\mathcal{L}[(.)] = \{H, (.)\}$ is the Liouvillian and $\{f, g\}$ are Poisson brackets defined according to

$$\{f, g\} := \sum_{i=1}^{N} \left(\frac{\partial f}{\partial X_i}\frac{\partial g}{\partial Q_i} - \frac{\partial f}{\partial Q_i}\frac{\partial g}{\partial X_i} \right). \tag{2.5}$$

An important quantity in statistical mechanics is the Shannon information entropy, which is given by

$$S(\varrho) = -\langle \ln \varrho \rangle, \tag{2.6}$$

which essentially quantifies the information content of the distribution ϱ. This interpretation follows since $S(\varrho)$ measures the amount of disorder in the state. For example, $S(\varrho)$ approaches zero as the distribution becomes a highly ordered state centred on a singular point in phase space, while $S(\varrho)$ approaches a maximum if ϱ is a highly disordered state spread evenly across all of the phase space. It is important to note that information is conserved within an isolated system [31], and using (2.4) we find that

$$\frac{d}{dt} S(\varrho(z_t;t)) = 0. \tag{2.7}$$

Another important information-theoretic quantity that we will frequently encounter is the Kullback-Leibler divergence. Considering two distributions $\varrho, \sigma \in \mathcal{P}$ with the support of ϱ contained in the support of σ, this divergence is given by

$$S(\varrho|\sigma) := \langle \ln \varrho \rangle - \langle \ln \sigma \rangle, \tag{2.8}$$

where the average is taken with respect to ϱ as before. According to Stein's lemma [46], the Kullback-Leibler divergence measures the difficulty of distinguishing samples from these two distributions. One can show that $S(\varrho|\sigma) \geq 0$ with equality if and only if $\varrho = \sigma$, which is a necessary and sufficient condition for $S(\varrho|\sigma)$ to define a mathematical divergence.

A central aspect of statistical mechanics is the notion of equilibrium. Suppose the state of the system is given by some arbitrary distribution $\varrho(z_0;\lambda)$ at time $t = 0$, with with average energy $U = \langle H(z_0;\lambda) \rangle$ and $\lambda_t = \lambda$ held fixed. A fundamental postulate of

statistical mechanics is that an isolated system in the long time limit will eventually relax to a micro-canonical state that is evenly distributed across all phase space points within this energy shell [2]:

$$\pi(z_t;\lambda) \propto \delta(U - H(z_t;\lambda)), \quad t \gg 0. \tag{2.9}$$

If we take the thermodynamic limit $N \to \infty$, then it can be rigorously proven that this state is equivalent to Gibbs canonical ensemble [47]:

$$\pi(z_t;\lambda) \simeq \frac{e^{-\beta H(z_t;\lambda)}}{Z(\lambda)}; \quad Z(\lambda) := \int dz_t \; e^{-\beta H(z_t;\lambda)}. \tag{2.10}$$

In this case $\beta = (k_b T)^{-1}$, with k_b the Boltzmann constant and T the temperature. Throughout this thesis we will assume that $k_b = 1$ to simplify notation. This temperature quantifies how the entropy of the equilibrium distribution changes with energy:

$$\frac{\partial S}{\partial U} = \frac{1}{T}. \tag{2.11}$$

A key property about the equilibrium distribution (2.10) is that the energy and entropy can be specified purely in terms of the temperature T and control parameter λ. To make this explicit we denote $S^{eq}(\lambda, T) = -\langle \ln \pi(z_t;\lambda)\rangle^{eq}$ and $U^{eq}(\lambda, T) = \langle H(z_t;\lambda)\rangle^{eq}$ as the equilibrium entropy and energy respectively. Another important thermodynamic quantity is the free energy, which is defined according to

$$F(\varrho) = \langle H(z_t;\lambda)\rangle - TS(\varrho). \tag{2.12}$$

Denoting the corresponding equilibrium free energy by $F^{eq}(\lambda, T) = U^{eq}(\lambda, T) - TS^{eq}(\lambda, T)$, it is straightforward to prove the following information-theoretic relation between the non-equilibrium free energy of some arbitrary state ϱ and its equilibrium counterpart [48]:

$$F(\varrho) = F^{eq}(\lambda, T) + TS(\varrho|\pi). \tag{2.13}$$

Due to the fact that $S(\varrho|\pi)$ is a divergence, we can see that the equilibrium state π minimises the free energy with respect to temperature T. Alternatively one can show that the state $\varrho = \pi$ maximises the Shannon entropy $S(\varrho)$ for a fixed average energy, implying that the Gibbs state is the state of maximum information within a given energy shell [49]. These relations are suggestive of a close connection between information theory and thermodynamics, and we will later see this connection more explicitly when deriving the laws of thermodynamics and fluctuation relations. Once we have introduced a distinction

between work and heat, it will be seen that the Shannon entropy quantifies a minimum bound on heat flow, whilst the change in free energy limits the amount of extractable work from a system. From now on we will collectively refer to energy, entropy and free energy as the *thermodynamic potentials* for the system of interest.

Thus far we have considered a large isolated system with arbitrary control over its Hamiltonian. However, in thermodynamics one is typically restricted to control over a smaller subsystem of the composite system. To differentiate between a controllable region, which we refer to as the system $\mathcal{S}$, and an uncontrollable region, referred to as the reservoir $\mathcal{R}$, we divide the composite Hamiltonian $H(z_t;\lambda) := H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)$ as follows:

$$H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda) := H_{\mathcal{S}}(x_t;\lambda) + H_{\mathcal{R}}(y_t) + \gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t), \tag{2.14}$$

where $z_t = (x_t, y_t)$ denotes the collective phase-space coordinate of $\mathcal{S}$ and $\mathcal{R}$ respectively at time t. In this model all control is placed on the internal Hamiltonian of $\mathcal{S}$, given by $H_{\mathcal{S}}(x_t;\lambda)$, in contrast to the uncontrolled reservoir Hamiltonian $H_{\mathcal{R}}(y_t)$. The interaction is mediated by the additional term $V_{\mathcal{S}\cup\mathcal{R}}(z_t)$, with the proportionality constant γ governing the strength of this interaction. To derive the thermodynamic potentials of $\mathcal{S}$, we first compute its state $\pi_{\mathcal{S}}(x_t;\lambda)$ by integrating out the reservoir degrees of freedom:

$$\pi_{\mathcal{S}}(x_t;\lambda) := \int dy_t \; \pi_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda), \tag{2.15}$$

where $\pi_{\mathcal{S}\cup\mathcal{R}}$ is the global Gibbs state with respect to $H_{\mathcal{S}\cup\mathcal{R}}$. At this stage it is useful to consider the Taylor expansion of $\pi_{\mathcal{S}\cup\mathcal{R}}$ with respect to the coupling γ:

$$\pi_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda) = \pi^0_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)\Big[1 + \beta\gamma\delta V_{\mathcal{S}\cup\mathcal{R}}(z_t)\Big] + \mathcal{O}(\gamma^2) \tag{2.16}$$

where

$$\pi^0_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda) = \frac{e^{-\beta(H_{\mathcal{S}}(x_t;\lambda)+H_{\mathcal{R}}(y_t))}}{Z^0_{\mathcal{S}\cup\mathcal{R}}(\lambda)}; \quad Z^0_{\mathcal{S}\cup\mathcal{R}}(\lambda) := \int dz_t \; e^{-\beta(H_{\mathcal{S}}(x_t;\lambda_t)+H_{\mathcal{R}}(y_t))}. \tag{2.17}$$

and

$$\delta V_{\mathcal{S}\cup\mathcal{R}}(z_t) = V_{\mathcal{S}\cup\mathcal{R}}(z_t) - \langle V_{\mathcal{S}\cup\mathcal{R}}(z_t)\rangle^0_{\mathcal{S}\cup\mathcal{R}}, \tag{2.18}$$

with $\langle(.)\rangle^0_{\mathcal{S}\cup\mathcal{R}}$ an ensemble average with respect to the uncoupled state $\pi^0_{\mathcal{S}\cup\mathcal{R}}$. Note that without loss of generality, we can assume that

$$\int dz_t V_{\mathcal{S}\cup\mathcal{R}}(z_t)\pi^0_{\mathcal{R}}(y_t) = 0. \tag{2.19}$$

This follows since we can always include the λ-independent contribution $\langle V_{\mathcal{S}\cup\mathcal{R}}\rangle^0_{\mathcal{R}}$ within the system Hamiltonian as an offset term, without effecting the thermodynamics of the system. Combining (2.15) with (2.16) and (2.19) yields

$$\pi_{\mathcal{S}}(x_t;\lambda) = \pi^0_{\mathcal{S}}(x_t;\lambda) + \mathcal{O}(\gamma^2), \tag{2.20}$$

with $\pi^0_{\mathcal{S}}(x_t;\lambda_t)$ the Gibbs state with respect to bare Hamiltonian $H_{\mathcal{S}}(x_t;\lambda_t)$. It should be clear from this expansion that $\pi_{\mathcal{S}}(x_t;\lambda_t) \simeq \pi^0_{\mathcal{S}}(x_t;\lambda_t)$ becomes an accurate approximation whenever $\gamma^2 \ll 1$. Due to the fact that interactions between $\mathcal{S}$ and $\mathcal{R}$ will typically take place along the surface area of $\mathcal{S}$, if $\mathcal{S}$ is large enough this area will be negligible compared with its enclosed volume [7]. In this situation the interaction Hamiltonian will become negligible, and hence at finite temperatures away from absolute zero one has $\gamma^2 \ll 1$. This is known as the *weak-coupling approximation*, and is a central assumption of statistical mechanics.

The thermodynamic potentials can be computed using the partition function of $\mathcal{S}$:

$$Z^0_{\mathcal{S}}(\lambda) := \int dx_t \; e^{-\beta H_{\mathcal{S}}(x_t;\lambda)}. \tag{2.21}$$

We find the following useful formulas for the free energy, internal energy and entropy:

$$\begin{aligned} F^{eq}_{\mathcal{S}}(\lambda,T) &:= -\beta^{-1}\ln Z^0_{\mathcal{S}}(\lambda), \\ U^{eq}_{\mathcal{S}}(\lambda,T) &:= -\partial_\beta \ln Z^0_{\mathcal{S}}(\lambda), \\ S^{eq}_{\mathcal{S}}(\lambda,T) &:= \beta[U^{eq}_{\mathcal{S}}(\lambda,T) - F^{eq}_{\mathcal{S}}(\lambda,T)]. \end{aligned} \tag{2.22}$$

Furthermore, since

$$Z^0_{\mathcal{S}}(\lambda) = \frac{Z^0_{\mathcal{S}\cup\mathcal{R}}(\lambda)}{Z^0_{\mathcal{R}}}, \tag{2.23}$$

we find the following extensivity relation between the thermodynamic potentials of $\mathcal{S}$ and $\mathcal{R}$:

$$\chi^{eq}_{\mathcal{S}\cup\mathcal{R}}(\lambda,T) = \chi^{eq}_{\mathcal{S}}(\lambda,T) + \chi^{eq}_{\mathcal{R}}(T), \tag{2.24}$$

with $\chi = \{F, S, U\}$. This demonstrates that the energies and entropies of the composite state $\mathcal{S}\cup\mathcal{R}$ split into separate contributions, due to the lack of correlation between the two constituents.

2.2 Stochastic thermodynamics and fluctuation relations

In the previous section we introduced a set of thermodynamic potentials that can be assigned to a system in equilibrium. In this section we will see how these quantities relate to the first and second law of thermodynamics. This connection can be established by looking at the fluctuations in energy of a system as it evolves away from equilibrium.

We consider a composite state in $\mathcal{S} \cup \mathcal{R}$ described by $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t; t) \in \mathcal{P}$ at time t. This evolution is governed by the Liouville equation (2.4), with the time dependence of the Hamiltonian $H_{\mathcal{S}\cup\mathcal{R}}(z_t; \lambda)$ in (2.14) originating from a change in the external parameter $\lambda = \lambda_t$. In the case of a weak coupling, we have seen in (3.41) that the internal energy of the system is given by the ensemble average of the bare Hamiltonian $H_{\mathcal{S}}(x_t; \lambda)$. At the level of an individual trajectory x_t evolving from some initial point x_0 up to point x_τ, the change in *fluctuating* internal energy $\Delta u_{\mathcal{S}}(z_\tau; \tau) = H_{\mathcal{S}}(x_\tau; \lambda_\tau) - H_{\mathcal{S}}(x_0; \lambda_0)$ can be expressed as [39]

$$\Delta u_{\mathcal{S}}(z_\tau; \tau) := \int_0^\tau dt\ \dot{\lambda} \frac{\partial H_{\mathcal{S}}}{\partial \lambda}(x_t; \lambda) + \int_0^\tau dt\ \dot{x} \frac{\partial H_{\mathcal{S}}}{\partial x}(x_t; \lambda). \tag{2.25}$$

We identify the first contribution as the fluctuating work done on $\mathcal{S}$:

$$w(z_\tau; \tau) := \int_0^\tau dt\ \dot{\lambda} \frac{\partial H_{\mathcal{S}}}{\partial \lambda}(x_t; \lambda). \tag{2.26}$$

This interpretation follows since we can equivalently express this term as

$$w(z_\tau; \tau) = H_{\mathcal{S}\cup\mathcal{R}}(z_\tau; \lambda_\tau) - H_{\mathcal{S}\cup\mathcal{R}}(z_0; \lambda_0), \tag{2.27}$$

which follows from straightforward partial differentiation. In other words, $w(z_t; t)$ defines the fluctuating energy change of the composite $\mathcal{S} \cup \mathcal{R}$, and hence represents the external energy used to change the system after connecting $\mathcal{S}$ and $\mathcal{R}$ and changing λ. Alternatively, we can interpret $-w(z_t; t)$ as the work done *by* $\mathcal{S}$, which is the useful energy that can be extracted from the process. By conservation of energy, any work put into $\mathcal{S} \cup \mathcal{R}$ that doesn't change the internal energy of $\mathcal{S}$ must be dissipated into $\mathcal{R}$, and so we define the fluctuating heat as the second term in (2.25):

$$q(z_\tau; \tau) := \int_0^\tau dt\ \dot{x} \frac{\partial H_{\mathcal{S}}}{\partial x}(x_t; \lambda). \tag{2.28}$$

Note that the fluctuating work $w(z_\tau; \tau) = w(z_t[z_0]; t)$, along with the heat and internal energy change, is implicitly written as a function of the initial phase space point z_0

because the evolution of system trajectory x_t depends on the deterministic evolution of the collective phase space for $\mathcal{S} \cup \mathcal{R}$ according to Hamilton's equations (2.1), denoted by the transformation $z_0 \to z_\tau[z_0]$. However, both the fluctuating heat and work can be determined by monitoring the system degrees of freedom alone along a specific trajectory. If one takes an average of (2.25) with respect to the initial ensemble $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_0; 0)$ we obtain the *first law of thermodynamics*:

$$\Delta U_{\mathcal{S}} = \langle w \rangle + \langle q \rangle, \tag{2.29}$$

where $\Delta U_{\mathcal{S}} = \langle H_{\mathcal{S}}(x_\tau; \lambda_\tau) \rangle - \langle H_{\mathcal{S}}(x_0; \lambda_0) \rangle$ is the average change in internal energy of $\mathcal{S}$.

The first law (2.29) is simply a statement about energy conservation, and tells us that any work done in changing the external parameter λ must be divided between $\mathcal{S}$ and $\mathcal{R}$, since we consider $\mathcal{S} \cup \mathcal{R}$ as energetically isolated. However, without additional assumptions we cannot make any further predictions about the changes in work and heat for a given process. It is only when we impose various equilibrium boundary conditions on the configuration of $\mathcal{S} \cup \mathcal{R}$ that one can derive additional constraints on the variables $\langle w \rangle$ and $\langle q \rangle$. In fact, we will see that boundary conditions also constrain the higher order fluctuations in work and heat, which in turn will lead us to the second law of thermodynamics.

Let us first consider a situation in which the composite configuration of $\mathcal{S} \cup \mathcal{R}$ is given by the equilibrium distribution $\pi_{\mathcal{S}\cup\mathcal{R}}(z_0; \lambda_0)$ at time $t = 0$. We further assume that the coupling $\gamma \ll 1$ is vanishingly small, so that we can approximate the reduced state of $\mathcal{S}$ by $\varrho_{\mathcal{S}}(x_0; 0) \simeq \pi^0_{\mathcal{S}}(x_0; \lambda_0)$ according to (2.20). In this limit we neglect even the first order contributions from the interaction, and hence the evolution of $\mathcal{S}$ is approximately isolated from $\mathcal{R}$ as we vary λ. In this sense one can write

$$\frac{\partial \varrho_{\mathcal{S}}}{\partial t} = \mathcal{L}_\lambda[\varrho_{\mathcal{S}}], \tag{2.30}$$

with $\mathcal{L}_\lambda[(.)] = \{H_{\mathcal{S}}(x_t; \lambda), (.)\}$ the Liouvillian of the bare system Hamiltonian. Using Hamilton's equations (2.1), one can verify that the fluctuating heat $q(z_\tau; \tau)$ in (2.25) vanishes, implying that all work is done on changing the internal energy of the system:

$$w(z_\tau; \tau) = w(x_\tau; \tau) = H_{\mathcal{S}}(x_\tau; \lambda_\tau) - H_{\mathcal{S}}(x_0; \lambda_0). \tag{2.31}$$

For a given trajectory $x_t[x_0]$, the probability distribution for fluctuating work is given by [18]

$$P(w) := \int dx_0 \; \delta\big[w - H_{\mathcal{S}}(x_\tau[x_0]; \lambda_\tau) + H_{\mathcal{S}}(x_0; \lambda_0)\big] \pi^0_{\mathcal{S}}(x_0; \lambda_0), \tag{2.32}$$

where the integral is carried out over all initial points in the phase space of $\mathcal{S}$. This distribution describes the statistics of work done on $\mathcal{S}$ as it is driven away from equilibrium, and the average work can be obtained from

$$\langle w \rangle = \int dw \; P(w) w. \tag{2.33}$$

After time $t = \tau$, the parameter λ_τ is held fixed and we let the system relax to the corresponding equilibrium configuration $\pi^0_{\mathcal{S}}(x_\tau; \lambda_\tau)$. During this thermalisation step, no work is done on $\mathcal{S}$.

Let us now compare this process with the work done along a hypothetical time-reversed process. In this case we vary the parameter according to $\lambda^*_t = \lambda_{\tau - t}$, and assume that $\mathcal{S}$ is initially in the equilibrium configuration $\pi^0_{\mathcal{S}}(x^*_0; \lambda^*_0)$ with respect to the time-reversed Hamiltonian $H_{\mathcal{S}}(x^*_0; \lambda^*_0)$. In this situation the corresponding distribution of work done is given by

$$P_R(w) := \int dx^*_0 \; \delta\big[w - H_{\mathcal{S}}(x^*_\tau[x^*_0]; \lambda^*_\tau) + H_{\mathcal{S}}(x^*_0; \lambda^*_0)\big] \pi^0_{\mathcal{S}}(x^*_0; \lambda^*_0). \tag{2.34}$$

We further assume that the Hamiltonian is an *even* function of momentum. Note that in a time-reversed process, all points in phase space have opposite momenta: $x^*_t = [\vec{X}_{\tau - t}, -\vec{Q}_{\tau - t}]$. Due to our assumption about the Hamiltonian, this implies that $H_{\mathcal{S}}(x^*_t; \lambda^*_t) = H_{\mathcal{S}}(x_{\tau - t}; \lambda_{\tau - t})$ and $\pi^0_{\mathcal{S}}(x^*_0; \lambda^*_0) = \pi^0_{\mathcal{S}}(x_\tau; \lambda_\tau)$. It then follows from Liouville's theorem that these two processes are connected according to the *Crooks fluctuation relation* [26]:

$$\frac{P(w)}{P_R(-w)} = e^{\beta(w - \Delta F^{eq}_{\mathcal{S}})}, \tag{2.35}$$

where $\Delta F^{eq}_{\mathcal{S}} = F^{eq}_{\mathcal{S}}(\lambda_\tau, T) - F^{eq}_{\mathcal{S}}(\lambda_0, T)$ is the difference in equilibrium free energy between the initial and final equilibrium distributions, as defined in (3.41). If we multiply both sides of (2.35) by $P_R(-w)$ and integrate over w, one obtains the *Jarzynski equality* [18]:

$$\langle e^{-\beta w} \rangle = e^{-\beta \Delta F^{eq}_{\mathcal{S}}}. \tag{2.36}$$

Finally, we invoke the Jensen inequality, which states that $\langle f(x) \rangle \geq f(\langle x \rangle)$ for a convex function $f(x)$. Since the exponential function is convex, it follows from (2.36) that

$$\langle w \rangle \geq \Delta F^{eq}_{\mathcal{S}}. \tag{2.37}$$

The pair of relations (2.35) and (2.36) constitute the so-called fluctuation relations. The Crooks relation (2.35) tells us that the probability of doing work w along the forward

protocol is exponentially favoured against the corresponding work extracted along the reverse protocol. This time-asymmetry is a fundamental consequence of the equilibration that occurs at the boundaries of the protocol. Secondly, the Jarzynski equality (2.36) places constraints on the higher order moments of work during the protocol, and allows us to connect non-equilibrium work fluctuations to the corresponding equilibrium difference in free energy. Experimentally, one can monitor the work in order to measure $\Delta F_{\mathcal{S}}^{eq}$ [4]. As a result of these constraints on the work fluctuations, we are lead to a statement about the second law of thermodynamics, namely *'the work done on a system during a thermodynamic process is bounded by its change in free energy*', as shown by (2.37). This solidifies the free energy as the thermodynamic potential constraining the partitioning of energy exchange between work and heat within the first law (2.29).

Thus far we have considered the stochastic thermodynamics of a system isolated from $\mathcal{R}$, and initially at equilibrium. However, this situation can been extended to describe systems that begin arbitrarily far from equilibrium and interact with $\mathcal{R}$. To obtain this extension, one needs to introduce the concept of *non-equilibrium entropy production*. Let us now assume a small but finite coupling strength satisfying $\gamma^2 \ll 1$ so as to justify the weak-coupling approximation as before. In this regime we can associate the bare Hamiltonian $H_{\mathcal{S}}(x_t;\lambda)$ with the internal energy of $\mathcal{S}$, and the definitions for fluctuating heat and work satisfy (2.28) and (2.26) respectively. It is assumed that at the start of the protocol, the initial state of $\mathcal{S}\cup\mathcal{R}$ is given by

$$\varrho_{\mathcal{S}\cup\mathcal{R}}(z_0;0) := \varrho_{\mathcal{S}}(x_0;0)\pi^0_{\mathcal{R}}(y_0), \tag{2.38}$$

with $\pi^0_{\mathcal{R}}(y_0)$ representing the Gibbs state of $\mathcal{R}$ at temperature T. The assumption of weak coupling implies that there are no initial correlations between $\mathcal{S}$ and $\mathcal{R}$ before the protocol. However, since the coupling is still taken to be finite the evolution of $\mathcal{S}$ is obtained from the following Liouville equation:

$$\frac{\partial \varrho_{\mathcal{S}}}{\partial t} = \int dy_t\ \mathcal{L}_\lambda[\varrho_{\mathcal{S}\cup\mathcal{R}}], \tag{2.39}$$

with $\mathcal{L}_\lambda[(.)] = \{H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda),(.)\}$. This equation is a starting point to deriving a Markovian master equation for the dynamics of $\mathcal{S}$, though we do not cover the details here. The quantity of interest here is the fluctuating entropy production up to time $t=\tau$:

$$\Sigma(z_\tau;\tau) := \Delta s_{\mathcal{S}}(x_\tau;\tau) - \beta q(z_\tau;\tau). \tag{2.40}$$

where $\Delta s(x_\tau;\tau) = s(x_\tau;\tau) - s(x_0;0)$ and

$$s_{\mathcal{S}}(x_t;t) := -\ln \varrho_{\mathcal{S}}(x_t;t), \tag{2.41}$$

is known as the surprisal, which quantifies the amount of information gained from sampling the phase space point x_t [46]. One can see that the average surprisal yields the Shannon information entropy $\langle s_{\mathcal{S}} \rangle = S_{\mathcal{S}}(\varrho_{\mathcal{S}})$ of $\mathcal{S}$. In other words, the average entropy production $\langle \Sigma \rangle = \Delta S_{\mathcal{S}} - \beta \langle q \rangle$ defines the sum of average information entropy change and heat, while the stochastic quantity (2.40) measures the fluctuations in surprisal and heat along the process. In analogy to (2.32), the statistics of entropy production are given by the following distribution:

$$P(\Sigma) = \int dz_0 \; \delta\big[\Sigma - \Sigma(z_\tau[z_0];\tau)\big]\, \varrho_{\mathcal{S}\cup\mathcal{R}}(z_0;0). \tag{2.42}$$

To derive the more general version of the Crooks relation (2.35), we consider an equivalent reverse protocol λ_t^* and assume that $H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)$ is an even function of momentum as before. In this reverse process, it is assumed that the final configuration of $\mathcal{S}$, given by $\varrho_{\mathcal{S}}(x_\tau;\tau)$, is disconnected from $\mathcal{R}$ at the end of the forward protocol. In addition, we further assume that this final distribution is an even function of momentum so that it is invariant under time-reversal. The initial configuration for the reverse protocol is then given by the uncorrelated state

$$\varrho_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0) := \varrho_{\mathcal{S}}(x_0^*;0)\pi_{\mathcal{R}}(y_0^*) = \varrho_{\mathcal{S}}(x_\tau;0)\pi_{\mathcal{R}}^0(y_\tau), \tag{2.43}$$

and the distribution for entropy production is given by

$$P_R(\Sigma) = \int dz_0^* \; \delta\big[\Sigma - \Sigma(z_\tau^*[z_0^*];\tau)\big]\, \varrho_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0). \tag{2.44}$$

Collecting all of these assumptions together and applying Liouville's theorem leads to the following detailed fluctuation relation [24]:

$$\frac{P(\Sigma)}{P_R(-\Sigma)} = e^{-\Sigma}. \tag{2.45}$$

This relation tells us that values of entropy produced along the forward protocol are exponentially favoured against the opposite decrease in entropy produced along the reverse protocol, hence revealing the time-asymmetry associated with the non-equilibrium process. Integrating (2.45) yields an integral fluctuation relation

$$\langle e^{-\Sigma} \rangle = 1, \tag{2.46}$$

which places constraints on the moments of entropy production and generalises the Jarzysnki equality (2.36). Finally, applying the Jensen inequality gives $\langle \Sigma \rangle \geq 0$, which

represents the non-equilibrium version of the second law. In this setting, any system evolving in or away from equilibrium, while in weak contact with a thermal reservoir, must produce a positive amount of entropy on average. Equivalently, we can rewrite this law as follows:

$$\langle q \rangle \geq -T\Delta S_{\mathcal{S}}. \tag{2.47}$$

The second law expressed in this form is known as the Clausius inequality. This relation reveals a fundamental connection between information and thermodynamics, known as Landauer's principle [**?**]. It implies that erasing the information content of $\mathcal{S}$ by the amount $-\Delta S_{\mathcal{S}}$ requires a minimum amount of heat $\langle q \rangle$ dissipated into $\mathcal{R}$ on average. We can view (2.46) as a generalisation of the Jarzynski equality to open systems, since if one assumes that the initial and final configurations of $\mathcal{S}$ are canonical, namely $\varrho_{\mathcal{S}}(x_0; 0) = \pi^0_{\mathcal{S}}(x_0; \lambda_0)$ and $\varrho_{\mathcal{S}}(x^*_0; 0) = \pi^0_{\mathcal{S}}(x_\tau; \lambda_\tau)$, then from the first law (2.25) we see that the entropy production reduces to $\Sigma = w - \Delta F^{eq}_{\mathcal{S}}$. Substituting this into (2.45) and (2.46) recovers the form of the Crooks and Jarzynski fluctuations that we previously demonstrated for closed systems.

2.3 Hamiltonian of mean force for strongly-coupled systems

In the previous section the fluctuation theorems for weakly-coupled classical systems for arbitrary non-equilibrium processes were presented. These in turn lead to stochastic formulations of the first and second laws of thermodynamics. However, ultimately the goal will be to extend these relations to the case where the assumption $\gamma^2 \ll 1$ is no longer valid. For microscopic systems, the surface area of $\mathcal{S}$ can become comparable to its bulk volume, implying a break down of the weak-coupling assumption. This of course does not imply that thermodynamics breaks down, but instead one must take a new approach when defining notions of energy and entropy. Before we can extend the fluctuation relations, one first needs to investigate how to attribute equilibrium potentials to a strongly-coupled system.

Let us recall the general expressions for the total Hamiltonian (2.14) and reduced state of $\mathcal{S}$ given by (2.15). Rather than applying any approximations, we can formally express the state of $\mathcal{S}$ as follows:

$$\pi_{\mathcal{S}}(x_t; \lambda) := \frac{e^{-\beta \tilde{H}_{\mathcal{S}}(x_t;\lambda)}}{\tilde{Z}_{\mathcal{S}}(\lambda)}; \qquad \tilde{Z}_{\mathcal{S}}(\lambda) = \int dx_t \; e^{-\beta \tilde{H}_{\mathcal{S}}(x_t;\lambda)}, \tag{2.48}$$

where

$$\tilde{H}_{\mathcal{S}}(x_t;\lambda) := H_{\mathcal{S}}(x_t;\lambda) - T\ln\langle e^{-\beta\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)}\rangle_{\mathcal{R}}^{eq}, \tag{2.49}$$

is the so-called *Hamiltonian of mean force* (HMF). Here we have defined

$$\langle f(z_t)\rangle_{\mathcal{R}}^{eq} := \int dy_t\ f(z_t)\pi_{\mathcal{R}}^0(y_t), \tag{2.50}$$

as an average over the reservoir degrees of freedom with respect to the uncoupled Gibbs state of $\mathcal{R}$. In this sense we can interpret (2.49) as an effective Hamiltonian of $\mathcal{S}$, with the additional term taking into account the effect of the interaction. Clearly this term vanishes in the case where $\gamma^2 \ll 1$ as we have seen before.

The HMF in turn defines a new partition function for the system given by $\tilde{Z}_{\mathcal{S}}(\lambda)$. Naturally we can use this function to compute the associated thermodynamic potentials. By using the statistical-mechanical formulas (3.41), we define the equilibrium free energy by:

$$\tilde{F}_{\mathcal{S}}^{eq}(\lambda,T) := -\beta^{-1}\ln\tilde{Z}_{\mathcal{S}}(\lambda), \tag{2.51}$$

Of course this definition of free energy requires further justification. Jarzynski has shown that the reversible work required to drive a system quasi-statically through a series of equilibrium states at strong-coupling is exactly given by the change in the free energy function (2.51) [39]. If we consider a change in the total Hamiltonian (2.14) Then the free energy change, denoted $\Delta\tilde{F}_{\mathcal{S}}$, can be expressed as

$$\Delta\tilde{F}_{\mathcal{S}}^{eq} = \int_0^\tau dt\ \dot{\lambda}\left\langle\frac{\partial H_{\mathcal{S}}}{\partial\lambda}\right\rangle_S^{\text{eq}}, \tag{2.52}$$

where $\langle(..)\rangle_\lambda^{\text{eq}}$ denotes the expectation value with respect to the state (2.48). As the process is quasi-static and only the system degrees of freedom are controlled, (2.52) is nothing but the integrated power over time, or equivalently the integrated rate of change of total energy. This means we can equate $\Delta\tilde{F}_{\mathcal{S}}^{eq} = \langle w\rangle$, where $\langle w\rangle$ is the work done on the system. Crucially this remains consistent with the usual thermodynamic definition of free energy, which quantifies the system's ability to perform work. Furthermore, since

$$\tilde{Z}_{\mathcal{S}}(\lambda) = \frac{Z_{\mathcal{S}\cup\mathcal{R}}^0(\lambda)}{Z_{\mathcal{R}}^0}, \tag{2.53}$$

we find the following extensivity relation between the free energies of $\mathcal{S}$ and $\mathcal{R}$:

$$\tilde{F}_{\mathcal{S}}^{eq}(\lambda,T) = F_{\mathcal{S}\cup\mathcal{R}}^{eq}(\lambda,T) - F_{\mathcal{R}}^{eq}(T), \tag{2.54}$$

Intuitively, (2.56) implies that the system free energy is the work required to the system in the strongly-coupled composite state of $\mathcal{S} \cup \mathcal{R}$.

Since the internal energy and entropy are related to the free energy through a Legendre transform, we further define these potentials as follows:

$$\begin{aligned} \tilde{U}_{\mathcal{S}}^{eq}(\lambda, T) &:= -\partial_\beta \ln \tilde{Z}_{\mathcal{S}}(\lambda), \\ \tilde{S}_{\mathcal{S}}^{eq}(\lambda, T) &:= \beta[\tilde{U}_{\mathcal{S}}^{eq}(\lambda, T) - \tilde{F}_{\mathcal{S}}^{eq}(\lambda, T)]. \end{aligned} \tag{2.55}$$

which additionally satisfy the extensivity relations

$$\tilde{\chi}_{\mathcal{S}}^{eq}(\lambda, T) = \chi_{\mathcal{S}\cup\mathcal{R}}^{eq}(\lambda, T) - \chi_{\mathcal{R}}^{eq}(T), \tag{2.56}$$

with $\chi = \{S, U\}$. This additivity implies that the presence of the interaction leaves the equilibrium properties of the reservoir unchanged. Instead, the influence of the interaction is attributed to the equilibrium properties of the system alone [9].

While the thermodynamic potentials are well defined at equilibrium, it is interesting to consider the corresponding fluctuating potentials that underlay the ensemble averages (2.51) and (2.55). This will be necessary to develop a generalisation of stochastic thermodynamics to the strong-coupling regime.

We first note that the equilibrium internal energy can be expressed as $\tilde{U}_{\mathcal{S}}^{eq}(\lambda, T) = \langle \partial_\beta \big[\beta \tilde{H}_{\mathcal{S}}(x_t; \lambda)\big] \rangle_{\mathcal{S}}^{eq}$ where $\langle (.) \rangle_{\mathcal{S}}^{eq}$ denotes an average with respect to the effective equilibrium state (2.48). Similarly one finds $\tilde{S}_{\mathcal{S}}^{eq}(\lambda, T) = -\langle \ln \pi_{\mathcal{S}} \rangle_{\mathcal{S}}^{eq} + \beta^2 \langle \partial_\beta \tilde{H}_{\mathcal{S}}(x_t; \lambda) \rangle_{\mathcal{S}}^{eq}$. These quantities can be interpreted as equilibrium averages over a set of *fluctuating* thermodynamic potentials appearing inside the brackets $\langle (.) \rangle_{\mathcal{S}}^{eq}$. We propose that the fluctuating potentials for internal energy, entropy and free energy for states arbitrarily far from equilibrium are given respectively by

$$\begin{aligned} \tilde{u}_{\mathcal{S}}(x_t; \lambda) &:= \partial_\beta \big[\beta \tilde{H}_{\mathcal{S}}(x_t; \lambda)\big], \\ \tilde{s}_{\mathcal{S}}(x_t; \lambda) &:= -\ln \, \varrho_s(x_t; t) + \beta^2 \partial_\beta \tilde{H}_{\mathcal{S}}(x_t; \lambda), \\ \tilde{f}_{\mathcal{S}}(x_t; \lambda) &:= \tilde{u}_{\mathcal{S}}(x_t; \lambda) - \beta^{-1} \tilde{s}_{\mathcal{S}}(x_t; \lambda). \end{aligned} \tag{2.57}$$

These functions account for the temperature dependence of the mean force Hamiltonian, give the averages (2.51) and (2.55), and reduce to the standard thermodynamic potentials used in stochastic thermodynamics in the limit of weak coupling. We will show that these generalised fluctuating potentials can be connected into a consistent thermodynamic framework. The average non-equilibrium internal energy will be denoted by $\tilde{U}_{\mathcal{S}}(\lambda; t) = \langle \tilde{u}_S(x_t; \lambda_t) \rangle_{\mathcal{S}}$, with $\langle (..) \rangle_{\mathcal{S}} = \int dx_t \;\; \varrho_{\mathcal{S}}(x_t; t)(..)$ now an average with respect to a general

non-equilibrium state of the system. Similarly the average entropy will be denoted by $\tilde{S}_S(\lambda;t) = \langle \tilde{s}_S(x_t;\lambda_t)\rangle_{\mathcal{S}}$ and average free energy by $\tilde{F}_{\mathcal{S}}(\lambda;t) = \langle \tilde{f}_{\mathcal{S}}(x_t;\lambda)\rangle_S$. From (2.57) one sees that the non-equilibrium entropy at strong coupling involves a contribution from the Gibbs-Shannon entropy alongside a second term $\beta^2\langle \partial_\beta \tilde{H}_{\mathcal{S}}(x_t;\lambda)\rangle_{\mathcal{S}}$ that has previously been identified as an intrinsic entropy in the context of small-scale molecular motors [23].

It is not obvious that these potentials should be well-defined and additive for any given distribution of $\mathcal{S}\cup\mathcal{R}$, unlike the equilibrium counterparts (2.51) and (2.55). However, let us consider a particular class $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t) \in \mathcal{D}_\beta$ of distributions defined by

$$\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t) = \varrho_{\mathcal{S}}(x_t;t)\pi_{\mathcal{R}}(y_t|x_t), \tag{2.58}$$

where we place no restriction on the system configuration and

$$\pi_{\mathcal{R}}(y_t|x_t) = \frac{\pi^0_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)}{\int dy_t\ \pi^0_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)}, \tag{2.59}$$

is the *equilibrium* conditional probability for reservoir micro-state y_t given a particular micro-state of the system x_t, obtained through application of Bayes' theorem. Because the system Hamiltonian cancels in the fraction in (4.41) the dependence on the control parameter λ cancels in the expression for $\pi_{\mathcal{R}}(y_t|x_t)$. The class of states $\mathcal{D}_\beta$ has previously been introduced in [37] and referred to as the *stationary preparation class*, which describes a conditional equilibrium state on the reservoir. In this case for any micro-state selected from the system the resulting conditional statistics of the reservoir are equivalent to that of the total canonical state (2.20). From a dynamical perspective, it can be shown that the set of states (4.41) naturally emerge if the thermalisation time-scale of $\mathcal{R}$ is much faster than that of $\mathcal{S}$, as shown by Strassberg and Esposito in [50]. Furthermore, in the weak-coupling limit $\gamma^2 \ll 1$ we have $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t) \simeq \varrho_{\mathcal{S}}(x_t;t)\pi^0_{\mathcal{R}}(y_t)$, in which case the system becomes uncorrelated with the thermal reservoir. For this class of states one still has a well defined notion of temperature attributed to a thermal environment. This is manifested by a *generalised* additive relationship between the thermodynamic potentials, which we state as a Theorem:

Theorem 2.3.1. *Let $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t) \in \mathcal{D}_\beta$, and denote $U_{\mathcal{S}\cup\mathcal{R}}(\lambda_t;t) = \langle H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)\rangle$ as the internal energy of $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t)$, $S_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) = -\langle ln\ \varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t)\rangle$ the Gibbs-Shannon entropy and $F_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) = U_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) - \beta^{-1}S_{\mathcal{S}\cup\mathcal{R}}(\lambda;t)$ as the free energy. Then the following additive property holds:*

$$\tilde{\chi}_{\mathcal{S}}(\lambda;t) = \chi_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) - \chi^{eq}_{\mathcal{R}}(T), \tag{2.60}$$

where $\chi \in \{F, S, U\}$.

Proof. To express $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t)\in\mathcal{D}_\beta$ in a more useful form we use the following identity [37];

$$\begin{aligned}\pi_{\mathcal{R}}(y_t|x_t) &= \frac{\pi^0_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)}{\int dx_t\ \pi^0_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)},\\ &= \frac{e^{-\beta(H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t))}}{\int dy_t\ e^{-\beta(H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t))}}.\end{aligned}\tag{2.61}$$

We now note that the non-equilibrium internal energy is given by $\tilde{U}_{\mathcal{S}}(\lambda;t)=\langle\partial_\beta\big[\beta\tilde{H}_{\mathcal{S}}(x_t;\lambda)\big]\rangle_{\mathcal{S}}$. To proceed we expand the fluctuating internal energy function $\tilde{u}_{\mathcal{S}}(x_t;\lambda)=\partial_\beta\big[\beta\tilde{H}_{\mathcal{S}}(x_t;\lambda)\big]$;

$$\begin{aligned}\tilde{u}_{\mathcal{S}}(x_t;\lambda) &= \partial_\beta\big[\beta\tilde{H}_{\mathcal{S}}(x_t;\lambda)\big],\\ &= H_{\mathcal{S}}(x_t;\lambda)-\frac{\partial_\beta\langle e^{-\beta\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)}\rangle^{eq}_{\mathcal{R}}}{\langle e^{-\beta\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)}\rangle^{eq}_{\mathcal{R}}},\\ &= H_{\mathcal{S}}(x_t;\lambda)+\frac{\int dy_t\ e^{-\beta(H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t))}[H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)]}{\int dy_t\ e^{-\beta(H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t))}}+\partial_\beta\big[e^{-\beta F^{eq}_{\mathcal{R}}}\big],\\ &= H_{\mathcal{S}}(x_t;\lambda)+\int dy_t\ \pi_{\mathcal{R}}(y_t|x_t)[H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)]-U^{eq}_{\mathcal{R}}(T).\end{aligned}\tag{2.62}$$

Averaging both sides (2.62) with respect to $\varrho_{\mathcal{S}}(x_t;t)$ gives

$$\begin{aligned}\tilde{U}_{\mathcal{S}}(\lambda;t) &= \int dz_t\ \varrho_{\mathcal{S}}(x_t;t)\pi_{\mathcal{R}}(y_t|x_t)[H_{\mathcal{S}}(x_t;\lambda)+H_{\mathcal{R}}(y_t)+\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)]-U^{eq}_{\mathcal{R}}(T),\\ &= U_{\mathcal{S}\cup\mathcal{R}}(\lambda;t)-U^{eq}_{\mathcal{R}}(T).\end{aligned}\tag{2.63}$$

Turning now to the entropy, we need to evaluate the Gibbs-Shannon entropy of the state $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t)\in\mathcal{D}_\beta$. This can be done from the following equivalent identity;

$$\pi_{\mathcal{R}}(y_t|x_t)=e^{-\beta(H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)-\tilde{H}_{\mathcal{S}}(x_t;\lambda)-F^{eq}_{\mathcal{R}})}.\tag{2.64}$$

Using this we can show the following

$$\begin{aligned}S_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) &= -\int dz_t\ \varrho_{\mathcal{S}}(x_t;t)\pi_{\mathcal{R}}(y_t|x_t)[\ln\ \varrho_{\mathcal{S}}(x_t;t)+\ln\ \pi_{\mathcal{R}}(y_t|x_t)],\\ &= S_{\mathcal{S}}(\lambda;t)-\beta F^{eq}_{\mathcal{R}}(T)+\beta\int dz_t\ \varrho_{\mathcal{S}}(x_t;t)\pi_{\mathcal{R}}(y_t|x_t)[H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)-\tilde{H}_{\mathcal{S}}(x_t;\lambda)],\\ &= S_{\mathcal{S}}(\lambda;t)-\beta(U_{\mathcal{S}\cup\mathcal{R}}(\lambda;t)-U^{eq}_{\mathcal{R}}(T))+S^{eq}_{\mathcal{R}}(T)-\beta\int dx_t\ \varrho_{\mathcal{S}}(x_t;t)\tilde{H}_{\mathcal{S}}(x_t;\lambda),\\ &= S_{\mathcal{S}}(\lambda;t)+\beta\tilde{U}_{\mathcal{S}}(\lambda;t)-\beta\langle\tilde{H}_{\mathcal{S}}(x_t;\lambda)\rangle_{\mathcal{S}}+S^{eq}_{\mathcal{R}}(T),\\ &= \tilde{S}_{\mathcal{S}}(\lambda;t)+S^{eq}_{\mathcal{R}}(T),\end{aligned}\tag{2.65}$$

where we used $U_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) - U^{eq}_{\mathcal{R}}(T) = \tilde{U}_{\mathcal{S}}(\lambda;t)$ and $\beta^2\langle\partial_\beta\tilde{H}_{\mathcal{S}}(x_t;\lambda)\rangle_{\mathcal{S}} = \beta\tilde{U}_{\mathcal{S}}(\lambda;t) - \beta\langle\tilde{H}_{\mathcal{S}}(x_t;\lambda)\rangle_{\mathcal{S}}$. Finally, the last additive relation

$$F_{\mathcal{S}\cup\mathcal{R}}(\lambda;t) = \tilde{F}_{\mathcal{S}}(\lambda;t) + F^{eq}_{\mathcal{R}}(T), \tag{2.66}$$

follows trivially from (2.62) and (2.65) together with the definition of fluctuating free energy, $\tilde{f}_{\mathcal{S}}(x_t;\lambda) = \tilde{u}_{\mathcal{S}}(x_t;\lambda) - \beta^{-1}\tilde{s}_{\mathcal{S}}(x_t;\lambda)$. This concludes the proof of (2.60). ■

In (2.60) the thermodynamic potentials $\chi^{eq}_{\mathcal{R}}(T)$ are equivalent to those of an isolated canonical reservoir. The relation (2.60) implies that the equilibrium properties of $\mathcal{R}$ remain unchanged relative to the arbitrary state of the system, even in the presence of correlations due to strong interaction. In other words, while the reservoir marginal of $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t)$ is not a canonical distribution, the effect of the interaction on the reservoir potentials is negligible. This is physically intuitive considering that the reservoir is macroscopic relative to the microscopic size of the system.

In the next section we will see that the fluctuating potentials (2.57) provide a suitable extension of stochastic thermodynamics beyond weak-coupling. According to (2.60) one can interpret the free energy $\tilde{F}_{\mathcal{S}}(\lambda,t)$ as the work required to quasi-statically couple the system to the reservoir, resulting in the composite state (4.41). Similarly, by energy conservation the term $\tilde{U}_{\mathcal{S}}(\lambda,t)$ represents a contribution to the total energy not stemming from the bare reservoir. However, when we consider the entropy $\tilde{S}_{\mathcal{S}}(\lambda,t)$, there is no longer a clear information-theoretic interpretation of this thermodynamic function. Interestingly, under some fairly weak assumptions about the structure of $\mathcal{R}$ we can recover a connection between $\tilde{S}_{\mathcal{S}}(\lambda,t)$ and information theory. This is demonstrated by the following lemma:

Lemma 2.3.1. *Let $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t) \in \mathcal{D}_\beta$, and denote the marginal distribution of $\mathcal{R}$ by $\varrho_{\mathcal{R}}(y_t;t) = \int dx_t\ \varrho_{\mathcal{S}\cup\mathcal{R}}(z_t;t)$. We also denote $S_{\mathcal{R}}(t)$ as the Gibbs-Shannon entropy of $\varrho_{\mathcal{R}}(y_t;t)$, and $\delta S_{\mathcal{R}} = S_{\mathcal{R}}(t) - S^{eq}_{\mathcal{R}}$ as the difference between the Gibbs-Shannon entropies of the marginal reservoir state and the corresponding bare reservoir. Under the assumption that $\delta S_{\mathcal{R}} \ll 1$, we find that the entropy of $\mathcal{S}$ is equivalent to*

$$\tilde{S}_{\mathcal{S}}(\lambda,t) = S(\mathcal{S}|\mathcal{R}), \tag{2.67}$$

where $S(\mathcal{S}|\mathcal{R}) = S_{\mathcal{S}\cup\mathcal{R}}(\lambda,t) - S_{\mathcal{R}}(t)$ is the conditional Shannon entropy.

Proof. A direct combination of $\delta S_{\mathcal{R}} \ll 1$ and (2.60) with the definition of $S(\mathcal{S}|\mathcal{R})$ gives the desired result. ■

We conclude from the above lemma that if the reservoir entropy is sufficiently unaffected by the presence of the interaction, then the entropy of $\mathcal{S}$ is approximately equal to the conditional Shannon entropy. This retains a connection between the strong-coupling definition of entropy and information theory, since the entropy $S(\mathcal{S}|\mathcal{R})$ quantifies the amount of information needed to describe the outcome of random variable of $\mathcal{S}$ conditioned on a known outcome for $\mathcal{R}$ [46]. This conditional dependence on the reservoir results from the correlations created by the non-negligible interaction. Note that while the exact conditions needed to ensure the approximation $\delta S_{\mathcal{R}} \ll 1$ have not been specified, from a phenomenological perspective one would expect this to hold considering that the reservoir is sufficiently macroscopic, thereby making its entropy insensitive to any coupling to the system. In this sense only the thermodynamic properties of $\mathcal{S}$ are expected to be influenced by the interaction. It remains an open problem to provide a more rigorous analysis behind this approximation.

2.4 Entropy production and time-asymmetry

Now that we have identified a set of physically relevant thermodynamic potentials for strongly-coupled systems, it is possible to develop a consistent generalisation of stochastic thermodynamics. As a first step one needs to introduce a definition of fluctuating heat. To do this we consider a general non-equilibrium process operating at an arbitrarily large coupling strength, which is realised over a time interval $[0, \tau]$ by varying the Hamiltonian through a parameter change $\lambda = \lambda(t)$, with initial configurations $\lambda(0) = \lambda_0$ and $\lambda(\tau) = \lambda_\tau$. We make two assumptions about this process:

(i) At initial time $t = 0$ the system-bath is in a conditional equilibrium state $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_0; 0) = \varrho_{\mathcal{S}}(x_0; 0)\pi_{\mathcal{R}}(y_0|x_0) \in \mathcal{D}_\beta$, with $\varrho_{\mathcal{S}}(x_0; 0)$ specifying an initial arbitrary state for the system.

(ii) The composite state $\mathcal{S} \cup \mathcal{R}$ undergoes closed evolution during the time interval $[0, \tau]$ governed by Liouville's equation

$$\frac{\partial}{\partial t}\, \varrho_{\mathcal{S}\cup\mathcal{R}}(z_t; t) = \mathcal{L}_\lambda\big[\varrho_{\mathcal{S}\cup\mathcal{R}}(z_t; t)\big], \tag{2.68}$$

where $\mathcal{L}_\lambda[(..)]$ is the corresponding Liouvillian resulting from the change in the Hamiltonian (2.14) over time. The resulting final state is specified by $\varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau; \tau)$ with final system configuration $\varrho_{\mathcal{S}}(x_\tau; \tau) = \int dy_\tau\ \varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau; \tau)$.

Assumption (i) is necessary in order to have a well-defined notion of both temperature and the Hamiltonian of mean force (2.49) prior to the non-equilibrium process. Assumption (ii) ensures that we account for all exchanges of heat and work between the system and the reservoir. No restrictions are imposed on the final configuration of the system, and we denote the transformation by $\varrho_{\mathcal{S}}(x_0;0) \to \varrho_{\mathcal{S}}(x_\tau;\tau)$. Following the approaches taken in [7, 37, 38] we can use the fluctuating potentials in (2.57) to define the fluctuating heat dissipated from the system into the reservoir up to time τ as

$$\tilde{q}(z_\tau;\tau) := \tilde{u}_{\mathcal{S}}(x_0;\lambda_0) - \tilde{u}_{\mathcal{S}}(x_\tau;\lambda_\tau) + \int_0^\tau dt\ \frac{\partial}{\partial t}\tilde{u}_{\mathcal{S}}(x_t;\lambda), \tag{2.69}$$

which represents the sum of work done during the process and the decrease in internal energy of the system, in accordance with the first law of thermodynamics. Note that $\tilde{q}(z_\tau;\tau) = \tilde{q}(z_\tau[z_0];\tau)$ is implicitly written as a function of the initial phase space point z_0 because the evolution of point x_t depends on the deterministic evolution of the collective phase space for the system and reservoir, denoted by the transformation $z_0 \to z_\tau[z_0]$. However, the RHS of (2.69) indicates that the heat can be determined by monitoring the system degrees of freedom alone along a specific trajectory. If we take into account the full evolution of $\mathcal{S}\cup\mathcal{R}$, it is straightforward to show that the average dissipated heat is given by

$$\langle\tilde{q}\rangle = U_{\mathcal{S}\cup\mathcal{R}}(\lambda_\tau;\tau) - \tilde{U}_{\mathcal{S}}(\lambda_\tau;\tau) - U_{\mathcal{R}}^{eq}, \tag{2.70}$$

which follows from (2.60) combined with initial condition (i), along with the fact that the integral in (2.69) is equivalent to the difference in total energy, $H_{\mathcal{S}\cup\mathcal{R}}(z_\tau,\lambda_\tau) - H_{\mathcal{S}\cup\mathcal{R}}(z_0,\lambda_0)$. This heat is non-zero because, unlike the initial state, the final state will not generally belong to the class $\mathcal{D}_\beta$ and so the additive relation (2.60) will not hold for the final state in general. We now introduce a definition of fluctuating entropy production as the sum of dissipated heat and change in the fluctuating entropy of the system;

$$\tilde{\Sigma}(z_\tau;\tau) := \tilde{s}_{\mathcal{S}}(x_\tau;\lambda_\tau) - \tilde{s}_{\mathcal{S}}(x_0;\lambda_0) + \beta\tilde{q}(z_\tau;\tau). \tag{2.71}$$

For the definition (2.71) to be a physically relevant candidate for entropy production then it should not be negative on average, in accordance with the second law of thermodynamics. This is shown to be true according to the following theorem:

Theorem 2.4.1. *For a non-equilibrium process satisfying assumptions (i) and (ii), the average entropy production up to time τ is given by*

$$\langle\tilde{\Sigma}\rangle = S[\varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)||\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)], \tag{2.72}$$

where $S(\rho||\sigma) = \langle \ln \rho \rangle - \langle \ln \sigma \rangle$ *is the Kullback-Leibler divergence between the final configuration of* $\mathcal{S} \cup \mathcal{R}$ *and the corresponding conditional equilibrium state* $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) \in \mathcal{D}_\beta$*, given by*

$$\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) = \varrho_{\mathcal{S}}(x_\tau;\tau)\pi_{\mathcal{R}}(y_\tau|x_\tau). \tag{2.73}$$

Proof. We begin by expressing the decrease in non-equilibrium entropy for the NEQ process specified by assumptions (i) and (ii) as follows;

$$\begin{aligned}
-\Delta\tilde{S}_{\mathcal{S}} &= \tilde{S}_{\mathcal{S}}(\lambda_0;0) - \tilde{S}_{\mathcal{S}}(\lambda_\tau;\tau), \\
&= S_{\mathcal{S}\cup\mathcal{R}}(\lambda_0;0) - S^{eq}_{\mathcal{R}} - \tilde{S}_{\mathcal{S}}(\lambda_\tau;\tau), \\
&= S_{\mathcal{S}\cup\mathcal{R}}(\lambda_\tau;\tau) - S^{eq}_{\mathcal{R}} - S_{\mathcal{S}}(\lambda_\tau;\tau) - \beta^2\langle\partial_\beta\tilde{H}_{\mathcal{S}}(x_\tau;\lambda_\tau)\rangle_{\mathcal{S}}, \\
&= S_{\mathcal{S}\cup\mathcal{R}}(\lambda_\tau;\tau) - S^{eq}_{\mathcal{R}} - S_{\mathcal{S}}(\lambda_\tau;\tau) - \beta\tilde{U}_{\mathcal{S}}(\lambda_\tau;\tau) + \beta\langle\tilde{H}_{\mathcal{S}}(x_\tau;\lambda_\tau)\rangle_{\mathcal{S}},
\end{aligned} \tag{2.74}$$

where we recall $S_{\mathcal{S}}(\lambda;t) = \int dx_t \;\; \varrho_{\mathcal{S}}(x_t;t) \ln \;\; \varrho_{\mathcal{S}}(x_t;t)$ represents the Gibbs-Shannon entropy of the system. In the second line we applied the additivity of the non-equilibrium entropy, according to (2.60). This is ensured by our choice of initial conditions given by assumption (i). In the third line we used the fact that the Gibbs-Shannon entropy is invariant under closed evolution given by (2.68) [46]. The remaining steps follow from the definitions of $\tilde{S}_{\mathcal{S}}(\lambda;t)$ and $\tilde{U}_{\mathcal{S}}(\lambda;t)$.

Now we introduce the Kullback-Leibler divergence $S[\varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)||\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)]$ defined in (2.72). Using $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) = \varrho_{\mathcal{S}}(x_\tau;\tau)\pi_{\mathcal{R}}(y_\tau|x_\tau)$ according to (2.64), the KL divergence can be evaluated as follows;

$$\begin{aligned}
S[\varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)||\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)] &= \int dz_\tau \varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) \ln \;\; \left[\frac{\varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)}{\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)}\right], \\
&= -S_{\mathcal{S}\cup\mathcal{R}}(\lambda_\tau;\tau) + S_{\mathcal{S}}(\lambda_\tau;\tau) - \int dz_\tau \;\; \varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) \ln \;\; \pi_{\mathcal{R}}(y_\tau|x_\tau), \\
&= -S_{\mathcal{S}\cup\mathcal{R}}(\lambda_\tau;\tau) + S_{\mathcal{S}}(\lambda_\tau;\tau) - \beta F^{eq}_{\mathcal{R}} + \beta\langle H_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\lambda_\tau)\rangle_{\mathcal{S}\cup\mathcal{R}} \\
&\qquad - \beta\langle\tilde{H}_{\mathcal{S}}(x_\tau;\lambda_\tau)\rangle_{\mathcal{S}}, \\
&= \Delta\tilde{S}_{\mathcal{S}} + \beta[U_{\mathcal{S}\cup\mathcal{R}}(\lambda_\tau;\tau) - \tilde{U}_{\mathcal{S}}(\lambda_\tau;\tau) - U^{eq}_{\mathcal{R}}], \\
&= \Delta\tilde{S}_{\mathcal{S}} + \beta\langle\tilde{q}\rangle,
\end{aligned} \tag{2.75}$$

where we used (2.74) and $F^{eq}_{\mathcal{R}} = U^{eq}_{\mathcal{R}} - \beta^{-1}S^{eq}_{\mathcal{R}}$ in the penultimate line and (2.70) in the final line. Finally, using the definition of the entropy production (2.71) completes the proof of (2.72). ■

By (2.72) and the positivity of the Kullback-Leibler divergence, one has $\langle\tilde{\Sigma}\rangle \geq 0$ as desired. as a result one obtains a form of the Clausius inequality valid for arbitrary coupling strengths which becomes

$$\langle\tilde{q}\rangle \geq -T\Delta\tilde{S}_{\mathcal{S}}. \tag{2.76}$$

This inequality represents the strong-coupling generalisation of (2.47). Perhaps surprisingly, the Clausius inequality derived here within the strong coupling regime suggests that the change in Gibbs-Shannon entropy is generally insufficient to bound the minimum heat dissipated into the bath during a non-equilibrium process.

According to Stein's lemma [46], the divergence appearing in (2.72) can be interpreted as a measure of distinguishability between the final distribution and the corresponding conditional equilibrium state $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) \in \mathcal{D}_\beta$. Thus the further the final state is driven away from the equivalent state from $\mathcal{D}_\beta$, the greater the amount of entropy production after the process. If the dynamics governed by (2.68) are such that $\mathcal{S}\cup\mathcal{R}$ remains in the corresponding conditional equilibrium state in $\mathcal{D}_\beta$, the bound in (2.76) can be saturated at any given time t. However, in this situation the dissipated heat and entropy change are simultaneously zero; $\langle\tilde{q}\rangle = -T\Delta\tilde{S}_{\mathcal{S}} = 0$. The expression (2.72) can be interpreted as a generalisation of a phenomenon known as *lag* encountered in closed/weakly-coupled thermodynamic systems [51]. The entropy production quantifies the extent to which the configuration of the system-bath *lags* behind a hypothetical quasi-static process in which the configuration remains in the evolving conditional equilibrium state, $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_t;t) \in \mathcal{D}_\beta$. Figure 2.1 illustrates this effect.

Result (2.72) is consistent with previously derived expressions for average entropy production when the weak-coupling limit is taken. If one assumes $\gamma^2 \ll 1$ then the Hamiltonian of mean force (2.49) reduces to the system Hamiltonian $H_{\mathcal{S}}(x_t;\lambda)$ independent of temperature. As expected the heat becomes $\langle\tilde{q}\rangle \simeq \Delta\langle H_{\mathcal{R}}\rangle_{\mathcal{R}}$, where $\langle H_{\mathcal{R}}\rangle_{\mathcal{R}}$ is the average energy of the reservoir Hamiltonian evaluated with respect to the configuration of $\mathcal{R}$ at time t. Secondly, this also means the entropy change reduces to the change in Gibbs-Shannon entropy $\tilde{S}_{\mathcal{S}}(\lambda_\tau;\tau) \approx S_{\mathcal{S}}(\lambda_\tau;\tau) = -\int dx_\tau\ \varrho_{\mathcal{S}}(x_\tau;\tau)\ln\ \varrho_{\mathcal{S}}(x_\tau;\tau)$. Finally, it can also be seen that the conditional equilibrium state $\sigma(z_t;t) \in \mathcal{D}_\beta$ reduces to a system state uncorrelated with the isolated canonical bath; $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) \simeq \varrho_{\mathcal{S}}(x_\tau;\tau)\pi^0_{\mathcal{R}}(y_\tau)$. By comparison with (2.72), we obtain the same equality derived in [41, 52] which is

$$\langle\tilde{\Sigma}\rangle \simeq S[\varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)||\varrho_{\mathcal{S}}(x_\tau;\tau)\pi^0_{\mathcal{R}}(y_\tau)]. \tag{2.77}$$

It should be noted that (2.77) was originally derived for quantum systems in [41, 52], though in the weak-coupling regime the result is entirely statistical-mechanical in nature

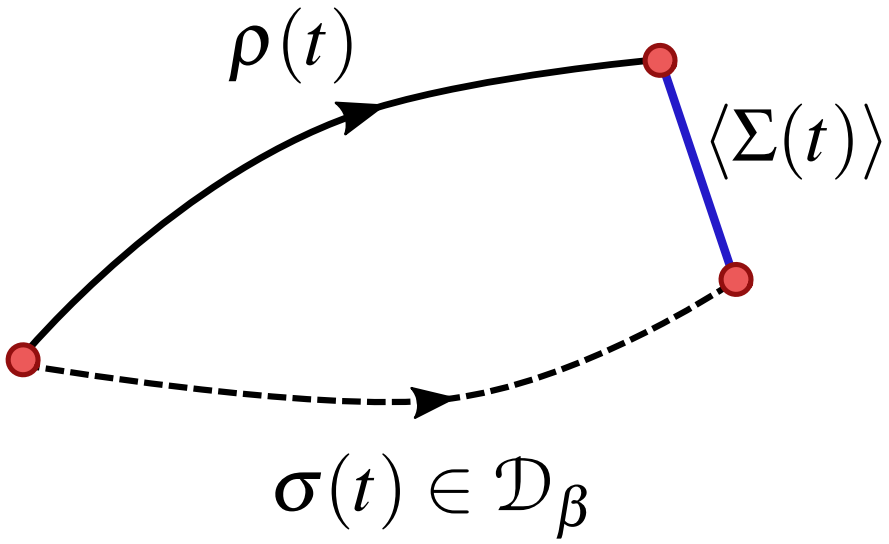


Figure 2.1: Schematic representation of the equality (2.72). The solid line represents the actual process given by the evolving distribution $\varrho(t) = \varrho_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau)$ whilst the dashed line represents a hypothetical quasi-static process in which the system-reservoir distribution stays in the conditional equilibrium state $\sigma(t) = \sigma_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\tau) \in \mathcal{D}_\beta$. The non-negative entropy production then quantifies the extent to which the system and bath are driven away from $\sigma(t)$, represented here as the distance of the blue line.

and continues to hold in classical systems. Finally, using the positivity of the KL divergence recovers the standard Clausius inequality (2.47) as expected.

We have demonstrated that the average entropy production $\langle\tilde{\Sigma}\rangle$ quantifies the extent to which the total system-reservoir is driven away from states in $\mathcal{D}_\beta$. This suggests that the fluctuations in entropy production can be used to quantify time-asymmetry in the dynamics of strongly coupled systems. In both weakly-coupled and closed systems, we have seen that the fluctuation relations indicate a breaking of time-reversal symmetry by comparing the statistics of positive entropy production for a forward trajectory versus negative entropy production along the corresponding time-reversed trajectory [26–28, 30, 31, 53]. We will now show that our generalised entropy production satisfies a Crooks-like fluctuation relation at arbitrarily strong couplings. Let us again suppose that we drive a system-reservoir configuration $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0) \in \mathcal{D}_\beta$ away from $\mathcal{D}_\beta$ by varying the control parameter λ_t over interval $t \in [0,\tau]$, and denote the initial and final configurations of the system by $\varrho_{\mathcal{S}}(x_0;0)$ and $\varrho_{\mathcal{S}}(x_\tau;\tau)$ respectively. The stochastic entropy production $\tilde{\Sigma}(z_\tau;\tau)$ along a particular phase space trajectory fluctuates according to the sampling of the initial phase space point, and the resulting probability of occurrence can be written

as follows;

$$P(\tilde{\Sigma}) = \int dz_0 \; \delta[\tilde{\Sigma} - \tilde{\Sigma}(z_\tau[z_0];\tau)]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0), \tag{2.78}$$

In complete analogy to the standard fluctuation relation (2.35), we will consider a corresponding time-reversed protocol $\lambda_t^* = \lambda_{\tau-t}$. As before, it is assumed that the total Hamiltonian is time-reversal symmetric, $H_{\mathcal{S}\cup\mathcal{R}}(z_t^*;\lambda_t) = H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda_t)$, where z_t^* indicates a time-reversed trajectory with a conjugated phase space point in which momentum is reversed. Secondly, the initial and final configurations of the system are also assumed to be time reversal symmetric; $\varrho_{\mathcal{S}}(x_0;0) = \varrho_{\mathcal{S}}(x_\tau^*;\tau)$ and $\varrho_{\mathcal{S}}(x_\tau;\tau) = \varrho_{\mathcal{S}}(x_0^*;0)$. For the time-reversed process, the initial configuration is given by $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0) = \varrho_{\mathcal{S}}(x_0^*;0)\pi_{\mathcal{R}}(y_0^*|x_0^*) \in \mathcal{D}_\beta$ and the control parameter is varied from $\lambda_t \to \lambda_0$. As with (2.78), entropy production along the reverse process has a corresponding probability of occurrence given by

$$P_R(\tilde{\Sigma}) = \int dz_0^* \; \delta[\tilde{\Sigma} - \tilde{\Sigma}(z_\tau^*[z_0^*];\tau)]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0). \tag{2.79}$$

The forwards and reverse probabilities are then connected via a fluctuation relation, which is stated in the following theorem:

Theorem 2.4.2. *Consider the entropy produced along the forward and backward protocols, with initial conditions sampled from the set of conditional equilibrium states $\mathcal{D}_\beta$. Then the entropy production along the forward process is exponentially favoured against the opposite amount along the reversed process:*

$$\frac{P(\tilde{\Sigma})}{P_R(-\tilde{\Sigma})} = e^{\tilde{\Sigma}}, \tag{2.80}$$

Proof. To begin, first note that the fluctuating heat (2.69) can be expressed in terms of the difference between the fluctuating total energy and fluctuating internal energy of the system;

$$\tilde{q}(z_\tau;\tau) = [H_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\lambda_\tau) - \tilde{u}_{\mathcal{S}}(x_\tau;\lambda_\tau)] - [H_{\mathcal{S}\cup\mathcal{R}}(z_0;\lambda_0) - \tilde{u}_{\mathcal{S}}(x_0;\lambda_0)]. \tag{2.81}$$

Recall that the initial state for the forward process is specified by $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0) = \varrho_{\mathcal{S}}(x_0;0)\pi_{\mathcal{R}}(y_0|x_0)$, whilst for the time-reversed process the initial configuration is given

by $\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0) = \varrho_{\mathcal{S}}(x_0^*;0)\pi_{\mathcal{R}}(y_0^*|x_0^*) \in \mathcal{D}_\beta$. Using (2.64) we expand the following;

$$\begin{aligned}
\ln\ \left[\frac{\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0)}{\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0)}\right] &= \ln\ \left[\frac{\varrho_{\mathcal{S}}(x_0;0)\pi_{\mathcal{R}}(y_0|x_0)}{\varrho_{\mathcal{S}}(x_0^*;0)\pi_{\mathcal{R}}(y_0^*|x_0^*)}\right], \\
&= \ln\ \left[\frac{\varrho_{\mathcal{S}}(x_0;0)}{\varrho_{\mathcal{S}}(x_0^*;0)}\right] - \beta\big[H_{\mathcal{S}\cup\mathcal{R}}(z_0;\lambda_0) - H_{\mathcal{S}\cup\mathcal{R}}(z_0^*;\lambda_0^*) \\
&\qquad\qquad - \tilde{H}_{\mathcal{S}}(x_0;\lambda_0) + \tilde{H}_{\mathcal{S}}(x_0^*;\lambda_0^*)\big], \\
&= \ln\ \left[\frac{\varrho_{\mathcal{S}}(x_0;0)}{\varrho_{\mathcal{S}}(x_\tau;\tau)}\right] - \beta\big[H_{\mathcal{S}\cup\mathcal{R}}(z_0;\lambda_0) - H_{\mathcal{S}\cup\mathcal{R}}(z_\tau;\lambda_\tau) \\
&\qquad\qquad - \tilde{H}_{\mathcal{S}}(x_0;\lambda_0) + \tilde{H}_{\mathcal{S}}(x_\tau;\lambda_\tau)\big], \\
&= \Delta\tilde{s}_{\mathcal{S}} + \beta\big[\Delta H_{\mathcal{S}\cup\mathcal{R}} - \Delta\tilde{H}_{\mathcal{S}} - \beta\Delta\partial_\beta\tilde{H}_{\mathcal{S}}\big], \\
&= \Delta\tilde{s}_{\mathcal{S}} + \beta\tilde{q}(z_\tau;\tau), \\
&= \tilde{\Sigma}(z_\tau[z_0];\tau), \qquad (2.82)
\end{aligned}$$

where we used the time-reversal symmetry assumptions for $H_{\mathcal{S}\cup\mathcal{R}}$ and $\varrho_{\mathcal{S}}$ and in the final line applied the definition (2.71). We now evaluate the reverse probability $P_R(-\tilde{\Sigma})$;

$$\begin{aligned}
P_R(-\tilde{\Sigma}) &= \int dz_0^*\ \delta[\tilde{\Sigma} + \tilde{\Sigma}(z_\tau^*[z_0^*];\tau)]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0), \\
&= \int dz_0\ \left|\frac{\partial z_0^*}{\partial z_0}\right|^{-1}\delta[\tilde{\Sigma} - \tilde{\Sigma}(z_\tau[z_0];\tau)]\left[\frac{\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0)}{\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0)}\right]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0), \\
&= \int dz_0\ \left|\frac{\partial z_\tau}{\partial z_0}\right|^{-1} e^{-\tilde{\Sigma}(z_\tau[z_0])}\delta[\tilde{\Sigma} - \tilde{\Sigma}(z_\tau[z_0];\tau)]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0), \\
&= \int dz_0\ e^{-\tilde{\Sigma}(z_\tau[z_0])}\delta[\tilde{\Sigma} - \tilde{\Sigma}(z_\tau[z_0];\tau)]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0), \\
&= e^{-\tilde{\Sigma}}\int dz_0\ \delta[\tilde{\Sigma} - \tilde{\Sigma}(z_\tau[z_0];\tau)]\sigma_{\mathcal{S}\cup\mathcal{R}}(z_0;0), \\
&= e^{-\tilde{\Sigma}}P(\tilde{\Sigma}), \qquad (2.83)
\end{aligned}$$

where in the second line we performed a change of variables $z_0^* \to z_0$ along with $\Sigma(z_\tau^*;\tau) = -\Sigma(z_\tau;\tau)$, in the third line we used (2.82), in the fourth line we used the fact that the Jacobian is equal to unity due to Liouville's theorem [2], and in the fifth line we pulled the exponential outside the integral due to the presence of the delta function. ■

The statement of Theorem 2.4.2 is the central result of this chapter, and represents the most general form of the detailed fluctuation relation valid for non-equilibrium states of $\mathcal{S}$, taking into account possible correlations with a thermal environment due to arbitrarily

strong interactions. Ultimately, the fluctuation relation provides a refined form of the second law, going beyond the usual inequality $\langle\tilde{\Sigma}\rangle \geq 0$. One can see from (2.80) that there is a clear asymmetry between the forwards and backwards statistics of entropy production. Taking the logarithm of both sides and performing an average over $P(\tilde{\Sigma})$ yields an alternative expression for the average entropy production:

$$\langle\tilde{\Sigma}\rangle = S\big[P(\tilde{\Sigma})||P_R(-\tilde{\Sigma})\big] \geq 0. \tag{2.84}$$

Following Stein's lemma again, we see that the average entropy production also quantifies the distinguishability between statistics of the forward and reverse non-equilibrium processes respectively. By comparison with (2.72), if the dynamics are such that the system and reservoir remain in their corresponding configuration in $\mathcal{D}_\beta$ then the LHS of (2.84) reduces to zero, implying the dynamics are completely symmetric in time as expected [30, 51]. This solidifies our interpretation of the entropy production (2.71) as a measure of time-asymmetry and irreversibility generalised to the strong coupling regime.

While we see from (2.84) that the entropy production must be positive on average, Theorem 2.4.2 tells us further information about the likelihood that a process can violate the second law in the single-shot regime. We exhibit this using the following lemma:

Lemma 2.4.1. *Consider a process satisfying the detailed fluctuation relation in Theorem 2.4.2. Then the probability of observing a negative amount of entropy production becomes exponentially suppressed according to*

$$P(\tilde{\Sigma} \leq -\varphi) \leq e^{-\varphi}. \tag{2.85}$$

Proof. Taking (2.80), we multiply both sides by $P_R(-\tilde{\Sigma})$ and integrate with respect to $\tilde{\Sigma}$. From the normalisation of $P_R(-\tilde{\Sigma})$ we get the integral fluctuation relation

$$\langle e^{-\tilde{\Sigma}}\rangle = 1. \tag{2.86}$$

We now recall the Chernoff bound for variable $a < 0$

$$P(\tilde{\Sigma} \leq -\varphi) \leq \langle e^{a\tilde{\Sigma}}\rangle \; e^{a\varphi}. \tag{2.87}$$

Setting $a = -1$ and combining this with the integral fluctuation relation completes the derivation. ■

By (2.85) one sees that the likelihood of observing negative entropy production becomes exponentially smaller the more negative the entropy becomes.

Another consequence of Theorem 2.4.2 is a fluctuation-dissipation relation generalised to the strong-coupling regime. We have seen that the strong-coupling version of the Clausius inequality (2.77) can only be saturated in the limit where the state of $\mathcal{S} \cup \mathcal{R}$ remains in a conditional equilibrium state in $\mathcal{D}_\beta$. This can be satisfied in the limit of an isothermal process, where the length of time taken to complete a process is taken to be infinity [37]. In the limit of a large but finite time interval, linear-response theory predicts that any excess heat dissipated into $\mathcal{R}$ should be proportional to the thermal fluctuations in the entropy production [54]. We find this to hold in the strong-coupling regime:

Lemma 2.4.2. *Let $P(\tilde{\Sigma})$ be a Gaussian distribution with mean $\langle\tilde{\Sigma}\rangle$ and variance $\langle\Delta\tilde{\Sigma}^2\rangle$. Furthermore, let $P(\tilde{\Sigma})$ satisfy Theorem 2.4.2. Then the dissipated heat satisfies the following fluctuation-dissipation relation:*

$$\langle\tilde{q}\rangle = \frac{T}{2}\langle\Delta\tilde{\Sigma}^2\rangle - T\Delta\tilde{S}. \tag{2.88}$$

Proof. For the random variable $\tilde{\Sigma}$, the cumulants are defined by

$$\kappa^k := (-1)^k \frac{d^k}{d\alpha^k} \ln\langle e^{-\alpha\tilde{\Sigma}}\rangle\big|_{\alpha=0}, \tag{2.89}$$

where $\kappa^1 = \langle\tilde{\Sigma}\rangle$ and $\kappa^2 = \langle\Delta\tilde{\Sigma}^2\rangle$. By assumption the distribution $P(\tilde{\Sigma})$ takes the Gaussian form

$$P(\tilde{\Sigma}) := \frac{1}{\sqrt{2\pi\langle\Delta\tilde{\Sigma}^2\rangle}} e^{-\frac{\tilde{\Sigma}-\langle\tilde{\Sigma}\rangle}{2\langle\Delta\tilde{\Sigma}^2\rangle}} \tag{2.90}$$

It is straightforward to show that any distribution of this form satisfies $\kappa^k = 0$ for $k \geq 3$. We next use the integral fluctuation relation $\langle e^{-\tilde{\Sigma}}\rangle = 1$, which follows from Theorem 2.4.2 as we saw before. Expanding this equality in terms of the cumulants, we get

$$\begin{aligned} 0 &= \ln\langle e^{-\tilde{\Sigma}}\rangle, \\ &= \sum_{k=1}^{\infty} \frac{(-1)^k}{k!}\kappa^k, \\ &= -\langle\tilde{\Sigma}\rangle + \frac{1}{2}\langle\Delta\tilde{\Sigma}^2\rangle, \end{aligned} \tag{2.91}$$

where in the final line we used $\kappa^k = 0$ for $k \geq 3$. Multiplying both sides by T and using the definition (2.40) completes the proof. ■

Note that Lemma 2.4.2 relies on the assumption that in the quasi-isothermal limit, the distribution $P(\tilde{\Sigma})$ becomes Gaussian. By treating $\mathcal{S} \cup \mathcal{R}$ as a closed system one can prove this rigorously from a perturbation expansion of the corresponding Fokker-Planck equation [55].

2.5 Volume, pressure and alternative approaches

In the formalism presented thus far, the volume of the system has been assumed to be fixed. However, by taking into account fluctuations in the volume, one can obtain an alternative interpretation of the mean force Hamiltonian in terms of enthalpy [7]. Let us first suppose that the state of $\mathcal{S} \cup \mathcal{R}$ is given by the isothermal-isobaric ensemble:

$$\pi^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda) := \frac{1}{\mathcal{Z}_{\mathcal{S}\cup\mathcal{R}}(\lambda)} e^{-\beta(H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)+P\mathscr{V}_{\mathcal{R}}(y_t))}. \tag{2.92}$$

Here we assume the Hamiltonian of $\mathcal{S} \cup \mathcal{R}$ is given as before by (2.14). In this situation one assumes that the volume of $\mathcal{R}$, denoted $\mathscr{V}_{\mathcal{R}}(y_t)$, fluctuates around equilibrium at fixed pressure P. This contrasts with the canonical ensemble (2.10), which has a *fixed* volume. As shown in [7], in the thermodynamic limit the canonical and isothermal-isobaric ensembles become indistinguishable due to ensemble equivalence, which removes any ambiguity in the particular choice of ensemble for $\mathcal{S} \cup \mathcal{R}$. After integrating over the reservoir degrees of freedom, the reduced state of $\mathcal{S}$ can be expressed as follows:

$$\pi^{\mathscr{V}}_{\mathcal{S}}(x_t;\lambda) = \frac{1}{\tilde{Z}^{\mathscr{V}}_{\mathcal{S}}(\lambda)} e^{-\beta(H_{\mathcal{S}}(x_t,\lambda)+\phi_{\mathcal{S}}(x_t))}, \tag{2.93}$$

where

$$\phi_{\mathcal{S}}(x_t) := -\frac{1}{\beta} \ln \langle e^{-\beta\gamma V_{\mathcal{S}\cup\mathcal{R}}(z_t)} \rangle^{\mathrm{eq}}_{\mathcal{R}}, \tag{2.94}$$

is referred to as the *solvation* Hamiltonian of mean force [7]. Here $\langle [..] \rangle^{\mathrm{eq}}_{\mathcal{R}} = \int dy_t \, \pi^{\mathscr{V}}_{\mathcal{R}}(y_t)[..]$ is now an average with respect to the *bare* isothermal-isobaric reservoir. Note that $\phi_{\mathcal{S}}(x_t)$ is independent of the external parameter λ, but implicitly depends on the pressure P of the surrounding reservoir.

Of course the volume of $\mathcal{S}$ also fluctuates at pressure P, and one can use $\phi_{\mathcal{S}}(x_t)$ to obtain a thermodynamic definition of its volume. However, the interesting point here is that there are two separate approaches to defining the volume, and this further leads to ambiguity in the definition of both the entropy and internal energy of $\mathcal{S}$. One approach, referred to in [7] as the *partial molar representation*, follows the same framework that

has been presented so far. In this case, the volume of $\mathcal{S}$ is defined from the solvation HMF according to

$$\tilde{v}_{\mathcal{S}}(x_t;\lambda) := \frac{\partial \phi_{\mathcal{S}}(x_t)}{\partial P}. \tag{2.95}$$

In this representation it can be shown that the average volume of $\mathcal{S}$ is equivalent to the difference between the total volume of $\mathcal{S}\cup\mathcal{R}$ minus the volume of the bare reservoir in the absence of any coupling to $\mathcal{S}$. This obeys the same additive form as the internal energy, entropy and free energy expressed in the additive relation (2.56) for the canonical ensemble. To be more specific, the average volume of $\mathcal{S}\cup\mathcal{R}$ is given by $\mathscr{V}_{\mathcal{S}\cup\mathcal{R}}(\lambda,T,P) = \int dz_t\ \pi^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)\mathscr{V}_{\mathcal{R}}(y_t)$, while the average volume of $\mathcal{S}$ is given by $\mathscr{V}_{\mathcal{S}}(\lambda,T,P) = \int dx_t\ \pi^{\mathscr{V}}_{\mathcal{S}}(x_t;\lambda)\tilde{v}_{\mathcal{S}}(x_t;\lambda)$. It can then be shown that

$$\mathscr{V}_{\mathcal{S}}(\lambda,T,P) = \mathscr{V}_{\mathcal{S}\cup\mathcal{R}}(\lambda,T,P) - \langle \mathscr{V}_{\mathcal{R}}(y_t)\rangle^{\mathrm{eq}}_{\mathcal{R}}. \tag{2.96}$$

As with (2.57), for an arbitrary system state $\varrho_{\mathcal{S}}(x_t;t)$ the fluctuating internal energy and entropy of $\mathcal{S}$ are defined by

$$\tilde{u}_{\mathcal{S}}(x_t,\lambda) := H_{\mathcal{S}}(x_t,\lambda) + \beta\frac{\partial}{\partial\beta}\phi_{\mathcal{S}}(x_t) - P\tilde{v}_{\mathcal{S}}(x_t), \tag{2.97}$$

$$\tilde{s}_{\mathcal{S}}(x_t,\lambda) := -\ln \varrho_{\mathcal{S}}(x_t;t) + \beta^2\frac{\partial}{\partial\beta}\phi_{\mathcal{S}}(x_t), \tag{2.98}$$

where we note that the internal energy now contains an additional pressure term. The fluctuating *enthalpy*, which is used to quantify the work done on $\mathcal{S}$ due to variations in both the parameter λ and changes in the volume, is subsequently given by

$$\tilde{h}_{\mathcal{S}}(x_t,\lambda) := \tilde{u}_{\mathcal{S}}(x_t,\lambda) + P\tilde{v}_{\mathcal{S}}(x_t). \tag{2.99}$$

If one consider variations in the parameter λ_t as a function of time $t\in[0,\tau]$, then the first law of thermodynamics at the stochastic level can be expressed as

$$\tilde{q}_{\mathcal{S}}(z_\tau,\tau) := \tilde{h}_{\mathcal{S}}(x_0;\lambda_0) - \tilde{h}_{\mathcal{S}}(x_\tau;\lambda_\tau) + \int_0^\tau dt\ \frac{\partial}{\partial t}\tilde{h}_{\mathcal{S}}(x_t;\lambda_t). \tag{2.100}$$

Here $\tilde{q}_{\mathcal{S}}(z_\tau,\tau)$ is the fluctuating heat dissipated into $\mathcal{R}$ up to time τ, which consists of the total decrease in enthalpy of $\mathcal{S}$ plus the work done along the process. The second term in (2.100) is the work done on $\mathcal{S}$, and is given by the integrated rate of change of enthalpy. The corresponding entropy production is again given by the change in stochastic entropy and heat, as in (2.71).

It is possible to generalise the fluctuation relation in Theorem 2.4.2 to account for fluctuations in volume. For this we introduce a class $\mathcal{D}_{\beta,P}$ of conditional equilibrium states of the form

$$\sigma^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_t;t) := \varrho_{\mathcal{S}}(x_t;t)\pi^{\mathscr{V}}_{\mathcal{R}}(y_t|x_t), \tag{2.101}$$

where

$$\pi^{\mathscr{V}}_{\mathcal{R}}(y_t|x_t) = \frac{\pi^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)}{\int dx_t\ \pi^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)}, \tag{2.102}$$

is the *equilibrium* conditional probability for reservoir micro-state y_t given a particular micro-state of the system x_t in the global isothermal-isobaric ensemble at temperature T and pressure P. For any non-equilibrium process we will again assume that the global Hamiltonian is time-reversal symmetric, and assume that the initial and final configurations of the system are time reversal symmetric; $\varrho_{\mathcal{S}}(x_0;0) = \varrho_{\mathcal{S}}(x^*_\tau;\tau)$ and $\varrho_{\mathcal{S}}(x_\tau;\tau) = \varrho_{\mathcal{S}}(x^*_0;0)$. We now state our generalised detailed fluctuation relation:

Theorem 2.5.1. *Suppose $\mathcal{S}\cup\mathcal{R}$ is driven under closed evolution from variations in parameter λ_t for $t\in[0,\tau]$. We suppose the initial configuration is given by $\sigma^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_0;0) := \varrho_{\mathcal{S}}(x_0;0)\pi^{\mathscr{V}}_{\mathcal{R}}(y_0|x_0) \in \mathcal{D}_{\beta,P}$ and denote $P(\tilde{\Sigma})$ as the distribution of entropy production. Here $\tilde{\Sigma} = \Delta\tilde{s} + \beta\tilde{q}$, where $\Delta\tilde{s}$ is given by (2.97) and $\tilde{q}$ is given by (2.100). We further consider a time-reversed process driven according to $\lambda^*_t = \lambda_{\tau-t}$, with the initial configuration is given by $\sigma_{\mathcal{S}\cup\mathcal{R}}(z^*_0;0) = \varrho_{\mathcal{S}}(x^*_0;0)\pi^{\mathscr{V}}_{\mathcal{R}}(y^*_0|x^*_0) \in \mathcal{D}_{\beta,P}$, and denote the distribution of entropy production by $P_R(\tilde{\Sigma})$. Then*

$$\frac{P(\tilde{\Sigma})}{P_R(-\tilde{\Sigma})} = e^{\tilde{\Sigma}}. \tag{2.103}$$

and hence

$$\langle\tilde{q}\rangle \geq -T\Delta\langle\tilde{s}\rangle. \tag{2.104}$$

Proof. The proof follows similar steps to that of Theorem 2.4.2. It is straightforward to show that

$$\pi^{\mathscr{V}}_{\mathcal{R}}(y_t|x_t) \propto e^{-\beta(H_{\mathcal{S}\cup\mathcal{R}}(z_t;\lambda)-H_{\mathcal{S}}(x_t;\lambda)-\phi_{\mathcal{S}}(x_t))}, \tag{2.105}$$

We next evaluate the following ratio:

$$\begin{aligned}
\ln\ \left[\frac{\sigma^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_0;0)}{\sigma^{\mathscr{V}}_{\mathcal{S}\cup\mathcal{R}}(z_0^*;0)}\right] &= \ln\ \left[\frac{\varrho_{\mathcal{S}}(x_0;0)\pi^{\mathscr{V}}_{\mathcal{R}}(y_0|x_0)}{\varrho_{\mathcal{S}}(x_0^*;0)\pi^{\mathscr{V}}_{\mathcal{R}}(y_0^*|x_0^*)}\right], \\
&= \ln\ \left[\frac{\varrho_{\mathcal{S}}(x_0;0)}{\varrho_{\mathcal{S}}(x_\tau;\tau)}\right] + \beta\big[\Delta H_{\mathcal{S}\cup\mathcal{R}} - \Delta H_{\mathcal{S}} - \Delta\phi_{\mathcal{S}}\big], \\
&= \Delta\tilde{s}_{\mathcal{S}} + \beta\big[\Delta H_{\mathcal{S}\cup\mathcal{R}} - \Delta\tilde{u}_{\mathcal{S}} - P\Delta\tilde{v}_{\mathcal{S}}\big], \\
&= \Delta\tilde{s}_{\mathcal{S}} + \beta\big[\Delta H_{\mathcal{S}\cup\mathcal{R}} - \Delta\tilde{h}_{\mathcal{S}}\big], \\
&= \Delta\tilde{s}_{\mathcal{S}} + \beta\tilde{q}(z_\tau;\tau), \\
&= \tilde{\Sigma}(z_\tau[z_0];\tau), \qquad (2.106)
\end{aligned}$$

The fluctuation relation then follows from the second part of the proof in Theorem 2.4.2. ■

The above fluctuation theorem represents the most general Crooks relation derived in the case where both energy and volume fluctuate for arbitrary transitions between non-equilibrium states of $\mathcal{S}$. If we consider canonical boundary conditions for $\mathcal{S}$, then Theorem 2.5.1 reduces to a less general detailed fluctuation relation derived by Jarzynski [7].

The above formalism reduces to that presented in the previous section if we classify the enthalpy as the internal energy (2.57). In this case Theorem 2.5.1 and Theorem 2.4.2 become equivalent. Notably, Jarzynski has shown that if the reservoir is macroscopic and one takes the thermodynamic limit, then ensemble equivalence implies that the reduced state of $\mathcal{S}$, namely (2.93) is equal to the reduced canonical state (2.15) [7]. Thus when taking the thermodynamic limit for $\mathcal{R}$ both forms of entropy production become the same quantity. This is important because it provides a useful interpretation for the correction term $\phi_{\mathcal{S}}(x_t)$ appearing in the Hamiltonian of mean force. For a sufficiently macroscopic reservoir, $\phi_{\mathcal{S}}(x_t)$ can be interpreted as an additional PV term representing fluctuations in the system volume due to the non-negligible interaction. These additional fluctuations then contribute to the enthalpy of the system. In contrast, in the weak-coupling limit all enthalpy changes are related only to fluctuations in the internal energy of the system, given by $H_{\mathcal{S}}(x_t;\lambda)$.

Intriguingly, an alternative representation of heat and entropy can also be conceived that also remains consistent with the first and second law of thermodynamics. We refer to this as the *bare* representation. Contrasting with the previous approach, one can instead

use the solvation HMF to define the volume of $\mathcal{S}$ according to

$$\tilde{v}'_{\mathcal{S}}(x_t) := \frac{\phi_{\mathcal{S}}(x_t)}{P}, \tag{2.107}$$

which is not equivalent to the volume (2.95) unless the interaction can be neglected. Furthermore, the internal energy can be equated to the bare Hamiltonian:

$$\tilde{u}'_{\mathcal{S}}(x_t, \lambda) := H_{\mathcal{S}}(x_t, \lambda), \tag{2.108}$$

and the entropy is simply the log-likelihood of the system's configuration:

$$\tilde{s}'_{\mathcal{S}}(x_t, \lambda) := -\ln \varrho_{\mathcal{S}}(x_t; \lambda), \tag{2.109}$$

which is equal to the Shannon information entropy on average. Taking these definitions together, the enthalpy of $\mathcal{S}$ in the bare representation becomes $\tilde{h}'_{\mathcal{S}}(x_t, \lambda) := \tilde{u}'_{\mathcal{S}}(x_t, \lambda) + P\tilde{v}'_{\mathcal{S}}(x_t)$. Finally, an alternative definition of heat, denoted $\tilde{q}'_{\mathcal{S}}(z_\tau, \tau)$, follows from replacement $\tilde{h}_{\mathcal{S}}(x_t, \lambda) \to \tilde{h}'_{\mathcal{S}}(x_t, \lambda)$ in (2.100). It can then be seen that these definitions of heat and entropy are also consistent with the second law of thermodynamics, namely

$$\langle \tilde{q}'_{\mathcal{S}} \rangle \geq -T\Delta\langle \tilde{s}'_{\mathcal{S}} \rangle. \tag{2.110}$$

The reason that this remain valid is because the rate of change of enthalpy in both the partial molar and bare representations are the same, thus the total work done on $\mathcal{S}$ remains unchanged in both frameworks. As a result, the total fluctuating entropy production is the same in both frameworks, and hence the Crook's relation in Theorem 2.5.1 remains invariant to these changes in definition.

It is important to note that both the partial-molar and bare representations become equivalent in the weak-coupling limit $\gamma^2 \ll 1$. Thus this apparent ambiguity in the thermodynamic formalism stems from the presence of non-negligible interactions between $\mathcal{S}$ and $\mathcal{R}$. As there is no clear energetic division between the system and reservoir, this leaves the definition of the system's volume somewhat ambiguous. While both representations lead to a consistent thermodynamic framework, there is currently no physical principle to discriminate between the two. It is further argued in [37] that while the notion of work is clearly well-defined regardless of the coupling γ, as work is always given by the total change in energy of $\mathcal{S} \cup \mathcal{R}$, the dissipated heat is an ambiguous quantity at both the ensemble and stochastic level. This again stems from the non-unique definition of the system's internal energy. This ambiguity does not modify the second law

of thermodynamics, since we have argued that entropy production remains the same in both frameworks. On the other hand, the partitioning between entropy and heat does in fact become arbitrary, and hence the particular splitting of the work into internal energy and heat within the first law amounts to a *gauge freedom*. This particular freedom is simply a matter of interpretation. If one adopts the partial molar representation, then the definitions of the thermodynamic potentials satisfy the usual extensivity relations and they are related via the usual Legendre transform [38]. On the other hand, within the bare representation the definition of heat provides a useful bound on the Shannon entropy change. If one is concerned with understanding fundamental bounds on the change in information contained within $\mathcal{S}$ in the strong-coupling regime, such as in Landauer erasure, then this approach is most relevant. It remains an open question as to whether or not a more solid definition of heat can be formulated in the strong-coupling regime.

2.6 Concluding remarks and outlook

In this chapter I have presented an overview of the thermodynamics of strongly-coupled classical systems. We have seen that the presence of a finite interaction between the system and its reservoir, which can occur when the system's volume is significantly small, imposes non-negligible modifications to the equilibrium properties of the system. In order to account for the coupling between $\mathcal{S}$ and $\mathcal{R}$, the Hamiltonian of mean force was introduced in order to describe the *effective* thermodynamics of $\mathcal{S}$ through a modified partition function. This allowed us to define effective forms of entropy, internal and free energy that remain consistent with the laws of thermodynamics, and this unified approach is able to successfully interpolate between the weak and strong-coupling regimes.

While the HMF is derived from standard equilibrium thermodynamics, I have shown that it can be applied to non-equilibrium processes. In order to apply the framework away from equilibrium I introduced a particular class of non-equilibrium states, whereby the system is prepared in an arbitrary state conditioned on an equilibrium reservoir. These conditional equilibrium states naturally emerge in the context of strongly-coupled systems where the thermalisation timescale of the reservoir is much faster than that of the system [50]. Using these states to account for correlations between $\mathcal{S}$ and $\mathcal{R}$, I derived an additive relation between the thermodynamic potentials, shown in Theorem 2.3.1. This result provides an intuitive interpretation of the generalised potentials (2.97) away from equilibrium, as it demonstrates that all system quantities can be computed by subtracting

the contributions from the macroscopic reservoir away from each thermodynamic potential. It was also shown in Lemma 2.3.1 that the resulting entropy of $\mathcal{S}$ is equivalent to the conditional Shannon entropy, thereby demonstrating that the thermodynamic entropy of $\mathcal{S}$ acquires additional contributions from the information contained in the correlations with $\mathcal{R}$.

In Section 2.4, through use of the effective internal energy obtained via the HMF, I established a formulation of the first law of thermodynamics at the stochastic level in (2.69) and proposed a definition of fluctuating heat valid at arbitrary coupling strengths. Combining this with the associated stochastic entropy change of the system, I defined a new notion of entropy production for strongly-coupled systems, extending the original definition first introduced by Crooks for weakly-coupled systems [26]. By using the Kullback-Leibler divergence, it was then proven in Theorem 2.4.1 that this entropy production quantifies the extent to which the composite state of $\mathcal{S} \cup \mathcal{R}$ is driven away from the corresponding conditional equilibrium state during a particular out-of-equilibrium process, and can only increase on average in accordance with the second law of thermodynamics. This result generalises a previous result obtained by Esposito *et al* in [41] by accounting for any initial correlations between $\mathcal{S}$ and $\mathcal{R}$ due to a non-negligible coupling. As a consequence of Theorem 2.4.1 we saw that a positive entropy production implies an excess amount of dissipated heat into $\mathcal{R}$, and hence establishes a strong-coupling version of the Clausius inequality (2.72).

In Theorem 2.4.2, I derived a generalised version of the detailed fluctuation relation with respect to this extended form of entropy production. This result demonstrates that positive entropy production is exponentially favoured against the corresponding amount along a hypothetical time-reversed process, in accordance with the standard Crooks relation in weakly-coupled systems [26]. It was then shown in Lemma 2.4.1 that the fluctuation provides a refinement of the second law, whereby large amounts of negative entropy produced along a non-equilibrium process are found to be exponentially unlikely. In addition to this, I showed in Lemma 2.4.2 that within the linear response regime, entropy production becomes proportional to the excess stochastic fluctuations. This represents a strong-coupling generalisation of the standard fluctuation-dissipation relation [18].

Finally, in Section 2.5 I presented an extension of this stochastic formalism to account for fluctuations in the volume of the system in strong contact with the reservoir. By taking into account the external pressure of $\mathcal{R}$, the HMF acquires a new interpretation as an additional PV term contributing to the changes in enthalpy of $\mathcal{S}$. As argued by Jarzynski

[7], ensemble equivalence implies that both approaches are essentially the same, differing only in the appropriate definition of the system's internal energy. However, by taking into account the pressure of $\mathcal{R}$, we can interpret the effect of the strong coupling as modifying the volume of the microscopic system, which is represented by the function (2.95). Using similar techniques to the previous formalism in Section 2.4, in Theorem 2.5.1 I derived an extended version of the detailed fluctuation relation with boundary conditions given by the conditional equilibrium states of an isothermal-isobaric reservoir. This extended the strong-coupling result obtained by Jarzynski in [7] to the case where the initial and final states of the system can be arbitrarily far from equilibrium. This result represents the most general classical detailed fluctuation relation, going beyond all previous theorems derived in the literature, and establishes the thermodynamic laws at arbitrary coupling strengths and state transitions. Intriguingly, it also shown in Section 2.5 that there are at least two definitions of heat that lead to the detailed fluctuation relation in Theorem 2.5.1. If one adopts a different definition of the system's volume (2.107), which is not equal to (2.95) unless the coupling is sufficiently weak, then a set of alternative thermodynamic potentials can be assigned to the system. In this bare representation, the internal energy of $\mathcal{S}$ is simply given by the bare Hamiltonian whilst the entropy is given by the Shannon information entropy. While these definitions lead to different forms of heat and entropy, the total stochastic entropy production remains invariant. This ambiguity is an inescapable consequence of the non-negligible coupling. One possible route to singling out a unique definition of heat may be obtainable by invoking the third law of thermodynamics. The third law predicts that the thermodynamic entropy of $\mathcal{S}$ should approach zero in the zero-temperature limit, which can be violated by the information entropy if correlations are present in the ground state of $\mathcal{S}$, and this is to be expected if the coupling term is finite. However, the entropy definition obtained from the partition function in the the partial molar representation, namely the average value of (2.97) at equilibrium, can be shown to vanish at absolute zero if $\mathcal{S}$ a Brownian particle [56]. This suggests that the approach adopted in this chapter is the only thermodynamic framework consistent with the third law of thermodynamics. If one is indeed concerned with ensuring the third law remains valid, then the form of heat given in (2.100) is singled out as the appropriate definition. Ultimately a more rigorous analysis of the third law for more general open systems will require quantum effects to be taken into account in the low-temperature regime.

Overall, the HMF proves to be a fundamental tool that can be used to describe thermodynamic processes in which the weak-coupling approximation breaks down. The

most pressing open question related to this work is how to generalise these results to the quantum regime. A notable obstacle to achieving this is the fact that the detailed fluctuation relations in Theorem 2.4.2 and Theorem 2.5.1 assume that the boundary conditions for $\mathcal{S}\cup\mathcal{R}$ are given by the class of conditional equilibrium states. Unfortunately, in quantum mechanics one cannot simply replace these states with a corresponding conditional density operator due to non-commutativity between the interaction Hamiltonian and the bare Hamiltonian of $\mathcal{S}$ and $\mathcal{R}$ [50]. Thus it is currently unclear how to derive a quantum detailed fluctuation relation for transitions between non-equilibrium states in the presence of strong coupling, and resolving this issues remains an important future goal.

CHAPTER 3

Temperature fluctuations in strongly-coupled quantum systems

It is known that temperature estimates of macroscopic systems in equilibrium are most precise when their energy fluctuations are large. The origin of this idea dates back to Bohr, who suggested that there should exist a form of complementarity between temperature and energy in thermodynamics similar to that of position and momentum in quantum theory [57]. His reasoning was that in order to assign a definite temperature T to a system it must be brought in contact with a thermal reservoir, in which case the energy U of the system fluctuates due to exchanges with the reservoir. On the other hand, to assign a sharp energy to the system it must be isolated from the reservoir, rendering the system's temperature T uncertain. Based on this heuristic argument Bohr conjectured the thermodynamic uncertainty relation:

$$\Delta\beta \geq \frac{1}{\Delta U}, \tag{3.1}$$

with $\beta = (k_{\mathrm{B}}T)^{-1}$ the inverse temperature. While (3.1) has since been derived in various settings [58–65], it was Mandelbrot who first based the concept of fluctuating temperature on the theory of statistical inference. Concretely, for a thermal system in canonical equilibrium, $\Delta\beta$ can be interpreted as the standard deviation associated with estimates of the parameter β. Mandelbrot proved that (3.1) sets the ultimate limit on simultaneous estimates of energy and temperature in classical statistical physics [58].

The notion of fluctuating temperature has proved to be fundamental in the emerging field of quantum thermometry, where advances in nanotechnology now allow temperature

sensing at sub-micron scales [66–78]. Using the tools of quantum metrology [79], the relation (3.1) can also be derived for weakly coupled quantum systems [67, 68, 70], where the equilibrium state is best described by the canonical ensemble. Within the grand-canonical ensemble the impact of the indistinguishability of quantum particles on the estimation of temperature and the chemical potential has also been explored [80]. Relation (3.1) informs us that when designing an accurate quantum thermometer one should search for systems with Hamiltonians that produce a large energy variance [70].

Recently there has been an emerging interest into the effects of strong coupling on temperature estimation [69, 71, 81]. As we have seen in the previous chapter, when moving below the nanoscale the strength of interactions between the system and the reservoir may become non-negligible, and the local equilibrium state of the system will not be of Gibbs form [13, 82]. In this regime thermodynamics needs to be adapted as the equilibrium properties of the system must now depend on the interaction energy [7, 9–12, 38–40, 44, 45, 50, 56, 83]. When one considers fully quantum systems, these modifications become more dramatic, as the potential non-commutativity between the Hamiltonian of the system and its interaction with the environment can lead to additional quantum coherences and entanglement. In this chapter I will present a quantum extension of the Hamiltonian of mean force that can be used to describe the equilibrium properties of strongly-coupled microscopic systems. It will be shown that the internal energy U of the system and its fluctuations ΔU are determined by a modified internal energy operator that differs from the bare Hamiltonian of the system [11]. Unlike the classical case, where all variables can be represented as continuous functions in phase space, this internal energy operator will not commute with the reduced density operator of the system in general. In this sense the strongly-coupled system can contain coherences in energy, and I will demonstrate that this phenomena directly influences its thermodynamic properties. In particular, it will be shown that coherences modify the equilibrium fluctuation-dissipation relation (FDR). In weakly-coupled systems, the FDR states that the heat capacity of the system is proportional to the variance in energy [2], which implies that the average energy of the system can only increase with increasing temperature. However, it will be shown that for strongly-coupled systems the FDR acquires non-classical modifications that are quantified by the average Wigner-Yanase-Dyson skew information [14, 84–86]. This is a quantity closely linked to measures of coherence, asymmetry and quantum speed limits [87, 88], and this this context measures the amount of additional energy coherence contained in the system density operator. Example studies of strongly-coupled open quantum systems have shown that the heat capacity can in

fact become negative at low temperatures [40, 56, 89, 90], implying that the energy can decrease with increasing temperature. This new FDR helps to explain the origin of this strong-coupling phenomenon, with negativity emerging due to the additional presence of quantum coherence.

This modification also brings into question the validity of (3.1) for general classical and quantum systems, and I investigate the impact of strong coupling on the thermodynamic uncertainty relation. By taking into account quantum properties of the effective internal energy operator and its temperature dependence, I here derive the general thermodynamic uncertainty principle valid at all coupling strengths. Formally this result follows from a general upper bound on the quantum Fisher information for exponential states. I prove that the quantum fluctuations arising from coherences between energy states of the system lead to increased fluctuations in the underlying temperature for a given spread in energy, and again relate to the Wigner-Yanase-Dyson skew information. This result is used to find a new upper bound on the achievable signal-to-noise ratio of an unbiased temperature estimate, and it identifies new constraints on the performance nanoscale thermometers that must be taken into account when probing the temperatures of strongly-coupled systems. These results are based on my published work "Energy-temperature uncertainty relation in quantum thermodynamics " [15].

In the final section of this chapter, I will present an exact analytical expression for the achievable accuracy of any quantum thermometer up to second order in coupling strength. This result provides a basis for investigating quantum thermometry setups beyond exactly-solvable models. The result is tested numerically on the spin-boson model, and I show that strong-coupling is detrimental to the performance of the quantum thermometer. Ultimately these result show that the known temperature uncertainty bounds are not necessarily achievable in the nanoscale regime in which the weak-coupling assumption breaks down.

3.1 Preliminaries: quantum states and observables

In this section I will outline the basic mathematical framework used to describe quantum systems. Let us denote $\mathcal{H}$ as a separable Hilbert space and $\mathcal{L}(\mathcal{H})$ the set of linear operators on this space. If the state of a quantum system is completely known then it can be represented by a wavefunction $\psi \in \mathcal{H}$, which is an element of the Hilbert space. More generally, if the state of the system undergoes a stochastic process then it can be described by an ensemble of wavefunctions $\{p_i, \psi_i\}_i$, with p_i the probability of the system being in

state ψ_i. Mathematically, this ensemble is described by a density operator, which belongs to the set of quantum states $\mathcal{S}(\mathcal{H})$ defined by

$$\mathcal{S}(\mathcal{H}) := \{\hat{\rho} \in \mathcal{L}(\mathcal{H}) \mid \hat{\rho} \geq 0,\ \mathrm{tr}[\hat{\rho}] = 1\}, \tag{3.2}$$

In general, a mixed state $\hat{\varrho} \in \mathcal{S}(\mathcal{H})$ is given by a convex mixture of pure states, namely

$$\hat{\varrho} = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{3.3}$$

As a result, any convex mixture of states $\hat{\rho}_1, \hat{\varrho}_2 \in \mathcal{S}(\mathcal{H})$ of the form $p\hat{\rho}_1 + (1-p)\hat{\rho}_2$ is also a state in $\mathcal{S}(\mathcal{H})$. The extremal points of the convex set $\mathcal{S}(\mathcal{H})$ satisfy $\hat{\rho}^2 = \hat{\rho}$ and represents the pure states $\hat{\rho} = |\psi\rangle\langle\psi|$.

The physical properties of a system are determined by the set of hermitian linear operators $\hat{A} \in \mathcal{L}(\mathcal{H})$ which satisfy $\hat{A}^\dagger = \hat{A}$. By the spectral theorem any function of a hermitian operator can be expressed $f(\hat{A}) = \sum_n f(a_n)\hat{P}_n$, where a_n are the eigenstates of the observable and $\hat{P}_n\hat{P}_m = \delta_{n,m}\hat{P}_n$ form a set of orthogonal projectors satisfying $\hat{P}_n^2 = \hat{P}_n$. It follows that the expectation value of $\hat{A}$ is given by

$$\langle\hat{A}\rangle := \mathrm{tr}[\hat{A}\hat{\rho}] = \sum_n a_n \mathrm{tr}[\hat{P}_n\hat{\rho}], \tag{3.4}$$

Here $\mathrm{tr}[\hat{P}_n\hat{\varrho}]$ represents the probability of observing the value a_n. While this outcome corresponds to a sharp measurement, one can consider more general kinds of measurement that are not sharp. A general quantum measurement is represented by a positive set of operators $\hat{\Pi}_x \geq 0$ that satisfy completeness $\int dx\ \hat{\Pi}_x = \hat{\mathbb{I}}$, where x represents a particular outcome and the integral is taken over the entire outcome space. The probability of observing outcome x is then given by the *Born rule*:

$$p(x) := \mathrm{tr}[\hat{\Pi}_x\hat{\rho}]. \tag{3.5}$$

Note that any quantum measurement has a non-unique decomposition $\hat{\Pi}_x = \hat{M}_x^\dagger\hat{M}_x$. It follows from the *measurement postulate* of quantum mechanics that the state after measurement is given by

$$\hat{\rho}' = \frac{\hat{M}_x^\dagger\ \hat{\varrho}\ \hat{M}_x}{p(x)}. \tag{3.6}$$

For projective measurements this implies that the state ends in an eigenstate of the corresponding observable outcome.

In a closed system, the dynamics of a state are determined by the Hamiltonian $\hat{H}(t)$ operator, which generally will depend on time. The evolving state of the system $\hat{\varrho}(t)$ then obeys the Liouville-von-Neumann equation:

$$\frac{d}{dt}\hat{\rho}(t) = -\frac{i}{\hbar}[\hat{H}(t), \hat{\rho}(t)], \tag{3.7}$$

If the initial state is given by $\hat{\rho}(0)$, then the general solution to (4.38) is given by $\hat{\varrho}(t) = \hat{U}(t,0)\hat{\varrho}(0)\hat{U}^{\dagger}(t,0)$, where

$$\hat{U}(t,0) := \overleftarrow{\exp}\bigg(-\frac{i}{\hbar}\int_0^t dt'\hat{H}(t')\bigg), \tag{3.8}$$

is the time-ordered evolution operator. If follows that this evolution is unitary so that $\hat{U}(t,0)\hat{U}^{\dagger}(t,0) = \hat{U}^{\dagger}(t,0)\hat{U}(t,0) = \hat{\mathbb{I}}$.

A composite Hilbert space may always be decomposed into a product of subsystems, such as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. If $\{|i\rangle_B\}$ is an orthonormal basis in $\mathcal{H}_B$, then the reduced state of A, denoted $\hat{\rho}_A$, is given by the partial trace:

$$\hat{\rho}_A = \mathrm{tr}_B[\hat{\rho}] = \sum_i \langle i|\,\hat{\rho}\,|i\rangle_B \in \mathcal{H}_A. \tag{3.9}$$

3.2 Quantum uncertainty and skew information

Our analysis throughout this chapter will rely on distinguishing between classical and non-classical fluctuations of observables in quantum mechanics, and I will present a framework for quantifying these different forms of statistical uncertainty for arbitrary mixed states. Consider a quantum state $\hat{\rho}$ and an observable $\hat{A}$. Wigner and Yanase considered the problem of quantifying the quantum uncertainty in observable $\hat{A}$ for the case where $\hat{\rho}$ is mixed [14]. However, they observed that the standard measure of uncertainty, namely the variance $\mathrm{Var}[\hat{\rho}, \hat{A}] := \mathrm{tr}[\hat{\rho}\,\delta\hat{A}^2]$ with $\delta\hat{A} = \hat{A} - \langle\hat{A}\rangle$, contains classical contributions due to mixing, and thus fails to fully quantify the non-classical fluctuations in the observable $\hat{A}$. This problem can be resolved by finding a quantum measure of uncertainty $Q[\hat{\rho}, \hat{A}]$ and classical measure $K[\hat{\rho}, \hat{A}]$ such that the variance can be partitioned according to

$$\mathrm{Var}[\hat{\rho}, \hat{A}] = Q[\hat{\rho}, \hat{A}] + K[\hat{\rho}, \hat{A}]. \tag{3.10}$$

Following the framework introduced by Luo [91], these functions are required to fulfil three conditions: (i) both terms should be non-negative, $Q[\hat{\rho}, \hat{A}] \geq 0$ and $K[\hat{\rho}, \hat{A}] \geq 0$, so that they can be interpreted as forms of statistical uncertainty, (ii) if the state $\hat{\rho}$ is

pure, then $Q[\hat{\rho}, \hat{A}] = \mathrm{Var}[\hat{\rho}, \hat{A}]$ while $K[\hat{\rho}, \hat{A}] = 0$ as all uncertainty should be associated to quantum fluctuations alone, while conversely $Q[\hat{\rho}, \hat{A}] = 0$ if $\hat{A}$ and $\hat{\rho}$ commute, (iii) $Q[\hat{\rho}, \hat{A}]$ must be convex with respect to $\hat{\rho}$, so that it decreases under classical mixing. Correspondingly, $K[\hat{\rho}, \hat{A}]$ must be concave with respect to $\hat{\rho}$.

The following function, known as the Wigner-Yanase-Dyson (WYD) skew information [14] was first shown to be a valid measure of quantum uncertainty satisfying conditions (i)-(iii):

$$Q_{1/2}[\hat{\rho}, \hat{A}] := -\frac{1}{2}\mathrm{tr}\big[[\hat{A}, \sqrt{\hat{\rho}}]^2\big]; \tag{3.11}$$

While conditions (i)-(iii) provide an interpretation of the skew information as a measure of statistical uncertainty, one can also use it to quantify the amount of coherence relative to observable $\hat{A}$ available in a given state $\hat{\rho}$. To see this we appeal to the *resource theory of asymmetry* [87, 92, 93]. To establish a measure of coherence, one needs to identify a set of states and operations that are considered 'free'. Free states, labelled by the class $\mathcal{S}_{free}(\mathcal{H})$ are those which are invariant under arbitrary rotation with respect to $\hat{A}$:

$$\hat{\rho}_{free} \in \mathcal{S}_{free}(\mathcal{H}) \Leftrightarrow e^{i\theta\hat{A}}\hat{\rho}_{free}e^{-i\theta\hat{A}} = \hat{\rho}_{free}; \quad \forall\theta \in \mathbb{R}. \tag{3.12}$$

In other words the resource under consideration here are the set of states diagonal in the basis of $\hat{A}$. Similarly, we can define the set of free operations $\mathcal{O}_{free}$ as the class of CPTP quantum maps ε_{free} that are invariant under arbitrary rotation with respect to $\hat{A}$ and any input state $\hat{\rho}$:

$$\varepsilon_{free}(.) \in \mathcal{O}_{free} \Leftrightarrow e^{i\theta\hat{A}}\varepsilon_{free}(\hat{\rho})e^{-i\theta\hat{A}} = \varepsilon_{free}\big(e^{i\theta\hat{A}}\hat{\rho}e^{-i\theta\hat{A}}\big); \quad \forall\theta \in \mathbb{R}. \tag{3.13}$$

To quantify the amount of coherence contained in a given state, one can now introduce a function from states to real numbers $f : \mathcal{S}(\mathcal{H}) \to \mathbb{R}$ such that $f(\varepsilon_{free}(\hat{\rho})) \leq f(\hat{\rho})$ for all free operations and $f(\hat{\rho}_{free}) = 0$ for all free states. In this sense the function f measures the amount of coherence in a given $\hat{\rho}$ relative to observable $\hat{A}$ since it vanishes for free states (ie. zero-resource states) and cannot be increased under free operations. It was subsequently proven in [87] that the skew information (3.12) satisfies these requirements as a valid measure of coherence in this resource-theoretic sense.

The skew information also plays an important role in deriving quantum speed limits [88, 94]. In particular, let us consider a quantum state $\hat{\rho}_t = \hat{U}_t\hat{\rho}_0\hat{U}_t^\dagger$ evolving in time $t \in [0, \tau]$, with the unitary evolution generated by a Hamiltonian $\hat{H}_t = -i\hbar\hat{U}_t\partial_t\hat{U}_t^\dagger$. It was subsequently proven in [88] that the skew information provides a lower bound on

the minimum time taken to evolve the state $\hat{\rho}_0 \to \hat{\rho}_\tau$, with the bound given by

$$\tau \geq \frac{\hbar}{\sqrt{2\langle Q_{1/2}[\hat{\rho}_t, \hat{H}_t]\rangle_\tau}} \mathcal{L}(\hat{\rho}_\tau, \hat{\rho}_0), \tag{3.14}$$

Here $\langle Q_{1/2}[\hat{\rho}_t, \hat{H}_t]\rangle_\tau$ represents the time-averaged skew information, which quantifies the amount of energy coherences maintained within the state during the evolution. The function $\mathcal{L}(\hat{\rho}_\tau, \hat{\rho}_0) = \arccos(\mathrm{tr}[\sqrt{\hat{\rho}_\tau}\sqrt{\hat{\rho}_0}])$ is known as the Hellinger angle, which represents a geometric distance over the state space $\mathcal{S}(\mathcal{H})$. The speed limit (3.14) hence provides a link between the skew information and the geometry of quantum states, and reveals that increasing the amount of quantum fluctuations in energy allows one to evolve a quantum state faster.

While (3.12) represents one particular way of quantifying quantum uncertainty and coherence, there is generally no unique way of separating the quantum and classical contributions to the variance. In fact, one can construct an entire family of different forms of skew information that all satisfy the same properties (i)-(iii) based on the class of monotone metrics over the quantum state space $\mathcal{S}(\mathcal{H})$ [95]. This is analogous to the family of different types of quantum entropy that act as equivalent measures of information [96].

However, certain forms of skew information can have physical relevance in particular settings. We now focus on a particular form of skew information first defined in [85, 86]:

$$Q[\hat{\rho}, \hat{A}] := \int_0^1 da\ Q_a[\hat{\rho}, \hat{A}], \tag{3.15}$$

where

$$Q_a[\hat{\rho}, \hat{A}] := -\frac{1}{2}\mathrm{tr}\big[[\hat{A}, \hat{\rho}^a][\hat{A}, \hat{\rho}^{1-a}]\big]; \quad a \in (0,1), \tag{3.16}$$

represents the quantum contribution with the complementary classical uncertainty given by

$$K[\hat{\rho}, \hat{A}] := \int_0^1 da\ K_a[\hat{\rho}, \hat{A}]. \tag{3.17}$$

and

$$K_a[\hat{\rho}, \hat{A}] := \mathrm{tr}\big[\hat{\rho}^a\, \delta\hat{A}\ \hat{\rho}^{1-a}\delta\hat{A}\big]; \quad a \in (0,1). \tag{3.18}$$

In order to verify that conditions (i)-(iii) are satisfied, the following useful representation will be used:

Lemma 3.2.1. *Let* $\hat{\rho} = \sum_n p_n |\psi_n\rangle \langle\psi_n|$ *denote a general mixed state of arbitrary rank. Then*

$$Q[\hat{\rho}, \hat{A}] = \sum_{n<m} \big(p_n + p_m - \Gamma(p_n, p_m)\big) \big| \langle\psi_n| \hat{A} |\psi_m\rangle \big|^2, \tag{3.19}$$

where

$$\Gamma(p_n, p_m) = \begin{cases} 0; & p_m p_n = 0, \\ 2p_n; & p_n = p_m \neq 0, \\ \frac{2(p_m - p_n)}{\ln p_m - \ln p_n}; & \textit{otherwise.} \end{cases} \tag{3.20}$$

Proof. For $a \in (0,1)$ the function $Q_a[\hat{\rho}, \hat{A}]$ can be expanded as

$$Q_a[\hat{\rho}, \hat{A}] = \sum_{n<m} \big(p_n + p_m - p_n^a p_m^{1-a} - p_m^a p_n^{1-a}\big) \big| \langle\psi_n| \hat{A} |\psi_m\rangle \big|^2, \tag{3.21}$$

Integrating with respect to a then completes the derivation. ■

Since $p_n + p_m \geq \Gamma(p_n, p_m)$ we have $Q[\hat{\rho}, \hat{A}] \geq 0$. Furthermore we have

$$\mathrm{Var}[\hat{\rho}, \hat{A}] = \sum_{n,m} \frac{p_n + p_m}{2} \big| \langle\psi_n| \delta\hat{A} |\psi_m\rangle \big|^2, \tag{3.22}$$

and using $p_n + p_m - \Gamma(p_n, p_m) \leq p_n + p_m$ gives $Q[\hat{\rho}, \hat{A}] \leq \mathrm{Var}[\hat{\rho}, \hat{A}]$. This implies $K[\hat{\rho}, \hat{A}] \geq 0$ as well, thus guaranteeing condition (i) holds. Condition (ii) is satisfied by taking $p_n = 1$ for $n = 1$ and $p_n = 0$ for $n > 1$. For condition (iii) we require Lieb's concavity theorem [97], which states that for any pair of positive-definite operators $\hat{X}, \hat{Y}$, the continuous mapping

$$(\hat{X}, \hat{Y}) \to \mathrm{tr}\big[\hat{Z}^\dagger \hat{X}^{1-a} \hat{Z} \hat{Y}^a\big]; \quad a \in [0,1], \tag{3.23}$$

is jointly concave in $(\hat{X}, \hat{Y})$, with $\hat{Z}$ a fixed linear operator. By comparison with (3.18), setting $\hat{X} = \hat{Y} = \hat{\rho}$ and $\hat{Z} = \delta\hat{A}$ proves that $K[\hat{\rho}, \hat{A}]$ is concave with respect to $\hat{\rho}$ due to the linearity of the integral over a in (3.17). As a result, the quantum uncertainty $Q[\hat{\rho}, \hat{A}]$ can only decrease under classical mixing as required by condition (iii).

Throughout the remainder of the chapter we will consider $Q[\hat{\rho}, \hat{A}]$ and $K[\hat{\rho}, \hat{A}]$ as the relevant measures of quantum and classical uncertainty, respectively. While this may appear to be an arbitrary choice, it will subsequently be shown that this form of skew information is intimately connected to thermodynamics.

3.3 Equilibrium states of strongly-coupled quantum systems

In this section I will discuss the equilibrium properties of quantum systems in contact with a large reservoir. Formally the distinction between system and reservoir is given by a decomposition of the Hilbert space into $\mathcal{H}_{\mathcal{S}\cup\mathcal{R}} = \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{R}}$. The energy and subsequent dynamics of $\mathcal{S} \cup \mathcal{R}$ is then governed by a generic Hamiltonian of the form

$$\hat{H}_{\mathcal{S}\cup\mathcal{R}} := \hat{H}_{\mathcal{S}} \otimes \hat{\mathbb{I}}_{\mathcal{R}} + \hat{\mathbb{I}}_{\mathcal{S}} \otimes \hat{H}_{\mathcal{R}} + \gamma \hat{V}_{\mathcal{S}\cup\mathcal{R}}, \tag{3.24}$$

where $\hat{H}_{\mathcal{R}}$ is the Hamiltonian of the reservoir $\mathcal{R}$, $\hat{V}_{\mathcal{S}\cup\mathcal{R}}$ is an arbitrary interaction term shared between $\mathcal{S}$ and $\mathcal{R}$ and γ is a scalar parameter quantifying the relative strength of the interaction. Formally if one considers $\mathcal{S} \cup \mathcal{R}$ to be isolated with a fixed total energy $U_{\mathcal{S}\cup\mathcal{R}}$ then the composite state is best described by the micro-canonical ensemble $\hat{\pi}_{\mathcal{S}\cup\mathcal{R}} \propto \delta(U_{\mathcal{S}\cup\mathcal{R}} - \hat{H}_{\mathcal{S}\cup\mathcal{R}})$, with inverse temperature $\beta = \partial S_{\mathcal{S}\cup\mathcal{R}}/\partial U_{\mathcal{S}\cup\mathcal{R}}$ and $S_{\mathcal{S}\cup\mathcal{R}}$ the thermodynamic entropy [98]. However, by imposing reasonable macroscopic conditions on the reservoir and taking the thermodynamic limit, we can safely replace the micro-canonical ensemble with the canonical one due to ensemble equivalence, so that $\hat{\pi}_{\mathcal{S}\cup\mathcal{R}} = e^{-\beta\hat{H}_{\mathcal{S}\cup\mathcal{R}}}/\mathcal{Z}_{\mathcal{S}\cup\mathcal{R}}$ with $\mathcal{Z}_{\mathcal{S}\cup\mathcal{R}} = \mathrm{tr}_{\mathcal{S}\cup\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{S}\cup\mathcal{R}}}]$ [47]. By taking the partial trace over $\mathcal{R}$, the state of the system at equilibrium is given by $\hat{\pi}_{\mathcal{S}} = \mathrm{tr}_{\mathcal{R}}[\hat{\pi}_{\mathcal{S}\cup\mathcal{R}}]$, which can be expressed in the following form:

$$\hat{\pi}_{\mathcal{S}} = \frac{e^{-\beta\hat{H}^*_{\mathcal{S}}}}{\mathcal{Z}^*_{\mathcal{S}}}; \qquad \mathcal{Z}^*_{\mathcal{S}} := \mathrm{tr}_{\mathcal{S}}[e^{-\beta\hat{H}^*_{\mathcal{S}}}], \tag{3.25}$$

where

$$\hat{H}^*_{\mathcal{S}} := -\frac{1}{\beta}\ln\left(\frac{\mathrm{tr}_{\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{S}\cup\mathcal{R}}}]}{\mathrm{tr}_{\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{R}}}]}\right), \tag{3.26}$$

is known as the quantum *Hamiltonian of mean force* [12, 39, 99]. Much like the classical counterpart considered in the previous chapter, this operator can be interpreted as an effective Hamiltonian describing $\mathcal{S}$. Unlike the bare Hamiltonian $\hat{H}_{\mathcal{S}}$, it implicitly depends on both the temperature T and interaction $\hat{V}_{\mathcal{S}\cup\mathcal{R}}$.

For a generic system-reservoir interaction model satisfying the weak-coupling approximation $\gamma^2 \ll 1$, I will subsequently show that $\hat{H}^*_{\mathcal{S}} \simeq \hat{H}_{\mathcal{S}}$, in which case the state of the system reduces to the familiar canonical ensemble, $\hat{\pi}_{\mathcal{S}} \simeq \hat{\pi}^0_{\mathcal{S}} = e^{-\beta\hat{H}_{\mathcal{S}}}/\mathcal{Z}_{\mathcal{S}}$ [2]. Furthermore, in this regime correlations between $\mathcal{S}$ and $\mathcal{R}$ can be neglected and the total state factorises into $\hat{\pi}_{\mathcal{S}\cup\mathcal{R}} \simeq \hat{\pi}^0_{\mathcal{S}} \otimes \hat{\pi}_{\mathcal{R}}$, where $\hat{\pi}_{\mathcal{R}} = e^{-\beta\hat{H}_{\mathcal{R}}}/\mathcal{Z}_{\mathcal{R}}$ and $\mathcal{Z}_{\mathcal{R}} = \mathrm{tr}_{\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{R}}}]$ is the

partition function for the reservoir in the absence of any coupling to $\mathcal{S}$ [35]. It is in this limit at which standard thermodynamics is formulated, and we can identify the energies of $\mathcal{S}$ and $\mathcal{R}$ with their bare Hamiltonians. This leads to a consistent notion of heat; any energy gained by $\mathcal{R}$ is the energy lost by $\mathcal{S}$ [35]. There are also notable cases where we can recover the usual Gibbs distribution even when γ is non-negligible. One example is if the interaction Hamiltonian is constructed from a random-matrix ensemble, in which case it can be rigorously proven that $\hat{H}_{\mathcal{S}}^{*} \simeq \hat{H}_{\mathcal{S}}$ regardless of the strength of interaction [100]. Heuristically, due to the random nature of the interaction its influence on $\mathcal{S}$ effectively averages out to zero. For spin and fermionic lattice systems, one can recover the local Gibbs state above a certain critical temperature, at which point correlations between the system and its surroundings decay exponentially [101].

However, in the general case where γ is non-negligible the system remains correlated with $\mathcal{R}$ and its thermodynamic properties are modified through the additional dependence on the coupling. It has been shown that the reduced Gibbs state (3.25) becomes the steady-state solution in more general non-linear system-reservoir interaction models, assuming an initial thermal reservoir uncorrelated with $\mathcal{S}$ [13]. In addition to explicit open system models, results based on typicality also demonstrate that the equilibrium state (3.25) emerges dynamically out of global pure state evolution for arbitrary interaction strengths [102]. Overall these results clearly motivate using (3.25) as the general state describing thermodynamic systems beyond weak coupling and canonical equilibrium. Before we discuss the thermodynamic role played by the HMF, it is worth analysing the dynamical models in which the state (3.25) emerges. The clearest example occurs in the paradigmatic model of Quantum Brownian Motion (QBM) [103–105]. In such a situation one wishes to construct a microscopic model for $\mathcal{S} \cup \mathcal{R}$ that reproduces a quantum Langevin equation, which is an equation that describes the motion of a damped oscillator in contact with a thermal environment. One particular model describing QBM, attributed to Caldeira and Leggett [103], consists of a single oscillator $\mathcal{S}$ linearly coupled to a collection of N harmonic oscillators that make up the reservoir $\mathcal{R}$, and the respective Hamiltonians in (2.14) are given by

$$\hat{H}_{\mathcal{S}} = \frac{\hat{p}^2}{2M} + \frac{M\omega^2\hat{x}^2}{2}, \qquad \hat{H}_{\mathcal{R}} := \sum_{j=1}^{N}\left(\frac{\hat{p}_j^2}{2M_j} + \frac{M_j\omega_j^2\hat{x}_j^2}{2}\right), \tag{3.27}$$

The interaction term is generically expressed in the linear form

$$\hat{V}_{\mathcal{S}\cup\mathcal{R}} := \sum_{j=1}^{N}\left(-\lambda_j\hat{x}\otimes\hat{x}_j + \frac{\lambda_j^2}{2M_j\omega_j^2}\hat{x}^2\right). \tag{3.28}$$

Here the position and momentum operators are given by $(\hat{x}, \hat{p})$ for $\mathcal{S}$ and $(\hat{x}_i, \hat{p}_i)$ for the i'th oscillator in $\mathcal{R}$, while $\gamma_i = \gamma C_i$ denotes the coupling constant between the system and each oscillator in the reservoir. Similarly (m, ω) and (m_i, ω_i) denote the mass and frequencies of the respective oscillators in $\mathcal{S}$ and $\mathcal{R}$. To accurately model the reservoir one takes the thermodynamic limit $N \to \infty$, and $\mathcal{R}$ becomes characterised by an Ohmic spectral density function. Examples of open quantum systems that can be modelled by (3.27) are nanoscale resonators, such as those manipulated in recent quantum control experiments [106], and also BEC impurities [107]. We remark here that one may construct alternative Hamiltonians that describe QBM, such as the continuum approach adopted in the Huttner-Barnett model, whose thermodynamic properties are explored in [44].

To allow a fully analytical solution, the reservoir frequencies are chosen equidistant, $\omega_j = j\Delta$ and the continuum limit is taken so that $\Delta \to 0$ (and $N \to \infty$). The coupling constants are chosen as the Drude-Ullersma spectrum [83],

$$\lambda_j = \sqrt{\frac{2\gamma M_j M \omega_j^2 \Delta}{\pi} \frac{\omega_{\mathrm{D}}^2}{\omega_{\mathrm{D}}^2 + \omega_j^2}}\,, \tag{3.29}$$

where ω_{D} is a large cutoff frequency.

Let us now suppose that prior to interactions the system and reservoir are uncoupled, so that the initial state at time $t = 0$ is of the form $\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(0) := \hat{\rho}_{\mathcal{S}}(0) \otimes \hat{\pi}_{\mathcal{R}}$, where $\hat{\rho}_{\mathcal{S}}(0)$ is an arbitrary state of $\mathcal{S}$ and the reservoir is assumed to be in canonical equilibrium with respect to $\hat{H}_{\mathcal{R}}$ at inverse temperature β. The benefit of the QBM model is that it is linear, and so the dynamics can be solved through an exact master equation governing the Wigner function of the system, regardless of the strength of the coupling γ [105]. Note that in the high temperature limit the model is equivalent to the classical Fokker-Planck process for a damped harmonic oscillator [104]. The exact master equation for the system dynamics is given by tracing out (4.38):

$$\frac{d}{dt}\hat{\rho}_{\mathcal{S}}(t) = -\frac{i}{\hbar}\mathrm{tr}_{\mathcal{R}}[[\hat{H}_{\mathcal{S}\cup\mathcal{R}}, \hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)]], \tag{3.30}$$

After taking the asymptotic time limit one finds that the steady state of the system is indeed given by the reduced canonical state (3.25), so that

$$\lim_{t\to\infty} \hat{\rho}_{\mathcal{S}}(t) = \mathrm{tr}_{\mathcal{R}}[\hat{\pi}_{\mathcal{S}\cup\mathcal{R}}] = \frac{e^{-\beta\hat{H}_{\mathcal{S}}^*}}{\mathcal{Z}_{\mathcal{S}}^*}. \tag{3.31}$$

The corresponding stationary Wigner function turns out to be of Gaussian form, and thus the resulting HMF for $\mathcal{S}$ takes the form of an effective harmonic oscillator parameterised

by a temperature-dependent mass and frequency [104]:

$$\hat{H}^*_{\mathcal{S}} = \frac{\hat{p}^2}{2M_T} + \frac{M_T\omega_T^2\hat{x}^2}{2} = \hbar\omega_T\left(\hat{n}_T + \frac{1}{2}\right), \tag{3.32}$$

where M_T and ω_T are given through the expectation values of $\hat{p}^2$ and $\hat{x}^2$ in the global thermal state, which are evaluated according to

$$M_T = \omega_T^{-1}\sqrt{\frac{\langle\hat{p}^2\rangle}{\langle\hat{x}^2\rangle}}, \tag{3.33}$$

$$\omega_T = 2T\ \mathrm{arcoth}(2\sqrt{\langle\hat{p}^2\rangle\langle\hat{x}^2\rangle}). \tag{3.34}$$

In its diagonal form the mean-force Hamiltonian contains a temperature-dependent number operator, $\hat{n}_T = \hat{a}_T^\dagger\hat{a}_T$, with annihilation operator $\hat{a}_T = \sqrt{\frac{A_T}{2\hbar}}\,(\hat{x} + \frac{i}{A_T}\,\hat{p})$ with $A_T = M_T\,\omega_T$. Using this we can diagonalise the state of $\mathcal{S}$ in terms of the number states of $\hat{n}_T$, so $\hat{\pi}_{\mathcal{S}} = e^{-\beta\hat{H}^*_{\mathcal{S}}}/\mathcal{Z}^*_{\mathcal{S}} = \sum_{n=0}^{\infty} p_n\,|n\rangle\,\langle n|$ where

$$p_n = \frac{e^{-\beta\epsilon_n}}{Z^*_{\mathcal{S}}}, \qquad \mathcal{Z}^*_{\mathcal{S}} = 2\,\sinh^{-1}\left(\frac{\beta\omega_T}{2}\right), \tag{3.35}$$

and $\epsilon_n = \omega_T(n + \frac{1}{2})$. In the continuum limit $N \to \infty$ the exact expressions for the quadratures are found to be [104]:

$$\langle\hat{x}^2\rangle = \frac{1}{M\beta\omega^2} + \frac{\hbar}{M\pi}\sum_{i=1}^{3}\left[\frac{(\nu_i - \omega_D)\Gamma(1 + \frac{\beta\hbar\nu_i}{2\pi})}{(\nu_{i+1} - \nu_i)(\nu_{i-1} - \nu_i)}\right], \tag{3.36}$$

$$\langle\hat{p}^2\rangle = M\omega^2\langle\hat{x}^2\rangle + \frac{\hbar M\gamma\omega_D}{\pi}\sum_{i=1}^{3}\left[\frac{\nu_i\Gamma(1 + \frac{\beta\hbar\nu_i}{2\pi})}{(\nu_{i+1} - \nu_i)(\nu_{i-1} - \nu_i)}\right], \tag{3.37}$$

where $\Gamma(z)$ is the digamma function and ν_i are the characteristic frequencies of the oscillator. In the limit of a large cutoff frequency, $\omega_D \gg \omega, \gamma$ the frequencies are given by

$$\begin{aligned}\nu_1 &= \frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} - \omega^2}, \\ \nu_2 &= \frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} - \omega^2}, \\ \nu_3 &= \omega_D - \gamma.\end{aligned} \tag{3.38}$$

Importantly, after taking the weak-coupling limit $\gamma \to 0$ one recovers the bare mass and frequency for $\mathcal{S}$ as expected. This exactly-solvable QBM model will be used throughout this chapter to analyse the thermodynamic properties of strongly-coupled quantum systems.

3.4 Heat capacity and the strong-coupling fluctuation-dissipation relation

As we saw in the previous chapter, in statistical mechanics the partition function encodes information about how the occupation probabilities are distributed between the different micro-states of $\mathcal{S}$, and it can be used to calculate the relevant thermodynamic potentials such as the free energy, internal energy and entropy. For a state in canonical equilibrium $\hat{\pi}_{\mathcal{S}}^0$, the free energy of $\mathcal{S}$ is given by $F_{\mathcal{S}} = U_{\mathcal{S}} - TS_{\mathcal{S}}$, with $U_{\mathcal{S}} = \mathrm{tr}_{\mathcal{S}}[\hat{\pi}_{\mathcal{S}}^0\ \hat{H}_{\mathcal{S}}]$ the average internal energy and $S_{\mathcal{S}} = -\mathrm{tr}_{\mathcal{S}}[\hat{\pi}_{\mathcal{S}}^0 \ln \hat{\pi}_{\mathcal{S}}^0]$ the von-Neumann entropy. From an information-theoretic perspective the canonical state represents the best possible guess for the configuration of $\mathcal{S}$ given prior knowledge of the system's energy, as it maximises the the information entropy given a fixed average energy according to the Maximum Entropy principle [49]. We can compute the free energy through knowledge of $\mathcal{Z}_{\mathcal{S}}$ since $F_{\mathcal{S}} := -\frac{1}{\beta}\ln \mathcal{Z}_{\mathcal{S}}$, and similar expressions follow for $U_{\mathcal{S}}$ and $S_{\mathcal{S}}$.

On the other hand, in the strong-coupling regime we obtain the alternative quantum partition function $\mathcal{Z}_{\mathcal{S}}^*$. As was done in the classical case, this motivates the following definition for the generalised free energy of the system for finite γ:

$$\tilde{F}_{\mathcal{S}} := -\frac{1}{\beta}\ln \mathcal{Z}_{\mathcal{S}}^*. \tag{3.39}$$

As we saw in the classical regime, this choice of free energy requires further justification. One may show that the reversible work required to drive a quantum system quasi-statically through a series of equilibrium states at strong-coupling is exactly given by the change in the free energy function (3.39) [99]. To see this let us parameterise the bare Hamiltonian of the system in (3.24) with some control parameter λ, so that $\hat{H}_{\mathcal{S}} = \hat{H}_{\mathcal{S}}(\lambda)$. For example, if our system is composed of a collection of spins then λ could denote the strength of an applied magnetic field. Now we assume that the parameter is smoothly varied in time between two values $\lambda_A \to \lambda_B$, such that the system always remains in the state (3.25) throughout the process whilst strongly interacting with $\mathcal{R}$. Then the free energy change, denoted $\Delta F_{\mathcal{S}} := F_{\mathcal{S}}(\lambda_B) - F_{\mathcal{S}}(\lambda_A)$, can be expressed as

$$\Delta\tilde{F}_{\mathcal{S}} = \int_A^B d\lambda \left\langle \frac{\partial \hat{H}_{\mathcal{S}}(\lambda)}{\partial \lambda} \right\rangle_{\lambda}^{\mathrm{eq}}, \tag{3.40}$$

where $\langle(..)\rangle_{\lambda}^{\mathrm{eq}}$ denotes the expectation value with respect to the state (3.25) at some fixed value of λ. As the process is quasi-static and only the system degrees of freedom are controlled, (3.40) is nothing but the integrated power over time, or equivalently

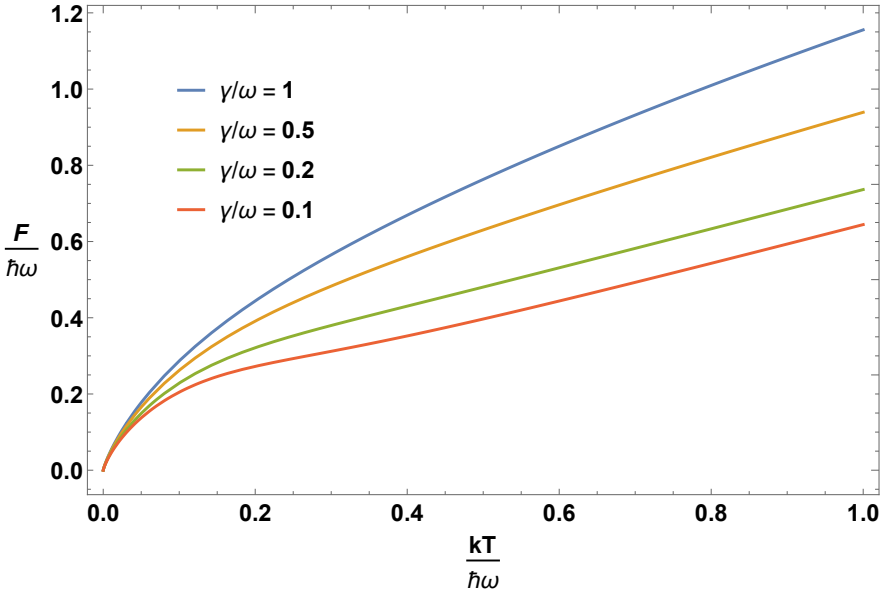


FIGURE 3.1. Free energy of the damped harmonic oscillator plotted as a function of temperature for different coupling strengths.

the integrated rate of change of total energy. This means we can equate $\Delta\tilde{F}_{\mathcal{S}} = \langle W \rangle$, where $\langle W \rangle$ is the work done on the system. Crucially this remains consistent with the usual thermodynamic definition of free energy, which quantifies the system's ability to perform work. As an example, Figure 3.1 demonstrates the free energy (3.39) as a function of temperature of the damped harmonic oscillator in the QBM model, which can be computed using (3.35). One can see that increasing the coupling also increases the overall free energy of $\mathcal{S}$ in the quantum regime. It can also be shown that in the infinite temperature limit, the free energy becomes independent of the coupling strength as expected.

With the definition of free energy established we can also obtain the respective thermodynamic entropy and internal energy of the strongly-coupled system, which are each related to the free energy according to

$$\tilde{S}_{\mathcal{S}} := \beta^2 \frac{\partial \tilde{F}_{\mathcal{S}}}{\partial \beta}, \qquad \tilde{U}_{\mathcal{S}} := \frac{\partial(\beta \tilde{F}_{\mathcal{S}})}{\partial \beta}. \tag{3.41}$$

We remark that these definitions satisfy the usual relation $\tilde{F}_{\mathcal{S}} = \tilde{U}_{\mathcal{S}} - T\tilde{S}_{\mathcal{S}}$ which means the thermodynamic potentials are related via the Legendre transform in the usual way. It can be shown that the partition function for $\mathcal{S}$ can be expressed as the ratio $\mathcal{Z}^*_{\mathcal{S}} = \mathcal{Z}_{\mathcal{S}\cup\mathcal{R}}/\mathcal{Z}_{\mathcal{R}}$,

which implies the following additive relation for the thermodynamic potentials [99]:

$$\tilde{\chi}_{\mathcal{S}} = \chi_{\mathcal{S}\cup\mathcal{R}} - \chi_{\mathcal{R}}, \tag{3.42}$$

where $\chi = F, S, U$. In other words, each potential for the system is equivalent to the difference between the potential for the total canonical state of $\mathcal{S}\cup\mathcal{R}$ and the isolated canonical state of $\mathcal{R}$. This additivity offers an intuitive interpretation of the expressions for $\tilde{F}_{\mathcal{S}}, \tilde{S}_{\mathcal{S}}, \tilde{U}_{\mathcal{S}}$. For example, the free energy $\tilde{F}_{\mathcal{S}}$ represents the reversible work required to immerse $\mathcal{S}$ into the composite state in $\mathcal{S}\cup\mathcal{R}$, which of course is given by the change in total free energy $F_{\mathcal{S}\cup\mathcal{R}} - F_{\mathcal{R}}$ expressed in (3.42) [7]. We remark here that this definition of internal energy leads to a consistent derivation of the Casimir effect in quantum electrodynamics, where the energy of the electromagnetic field is obtained by subtracting the contributions of the decoupled reservoir fields from the total energy [108].

Since the free energy, entropy and internal energies are defined through the partition function, they are essentially average quantities associated to the thermodynamic ensemble. However, one can also consider the corresponding operator expressions. For this we will make use of the following important operator identity [109]:

Lemma 3.4.1. *Let $\hat{A}(\theta) \in \mathcal{L}(\mathcal{H})$, where θ is some scalar parameter. Then the partial derivative of $e^{-\hat{A}(\theta)}$ with respect to θ is given by the following integral:*

$$\frac{\partial}{\partial\theta} e^{-\hat{A}(\theta)} := -\int_0^1 da\; e^{-(1-a)\hat{A}(\theta)} \partial_\theta[\hat{A}(\theta)] e^{-a\hat{A}(\theta)}, \tag{3.43}$$

Denoting expectation values by $\langle(..)\rangle^{\rm eq} = {\rm tr}_{\mathcal{S}}[(..)\hat{\pi}_{\mathcal{S}}]$ and using Lemma 3.4.1, the internal energy $\tilde{U}_{\mathcal{S}} = -\partial_\beta \ln \mathcal{Z}^*_{\mathcal{S}}$ can thus be rewritten as [15]:

$$\tilde{U}_{\mathcal{S}} = \langle \hat{E}^*_{\mathcal{S}} \rangle^{\rm eq}; \qquad \hat{E}^*_{\mathcal{S}} := \partial_\beta(\beta \hat{H}^*_{\mathcal{S}}). \tag{3.44}$$

Thus we can interpret $\hat{E}^*_{\mathcal{S}}$ as the internal energy operator for the strongly-coupled system. Note that this operator is not equivalent to $\hat{H}^*_{\mathcal{S}}$ due to the additional term $\partial_\beta \hat{H}^*_{\mathcal{S}}$ in (3.44). It is only in exceptional cases where the HMF becomes temperature-independent, such as in the bilinear coupling model for Brownian motion formulated in [110], that one has $\hat{E}^*_{\mathcal{S}} \simeq \hat{H}^*_{\mathcal{S}}$. In general this in-equivalence has interesting quantum-mechanical consequences, as it implies that the temperature-dependence of the HMF is responsible for the presence of energy coherences in the equilibrium state [15]. This follows from the fact that in general, one has

$$[\hat{E}^*_{\mathcal{S}}, \hat{\pi}_{\mathcal{S}}] \neq 0, \tag{3.45}$$

outside of the weak-coupling limit, leading to additional quantum uncertainty in the energy.

The specific heat capacity $\tilde{C}_{\mathcal{S}}$ of a thermal system at constant volume quantifies the susceptibility of the system to change its energy with respect to variations in temperature, and is formally defined as

$$\tilde{C}_{\mathcal{S}} := \frac{\partial \tilde{U}_{\mathcal{S}}}{\partial T}. \tag{3.46}$$

In the weak-coupling limit it is well known that $\tilde{C}_{\mathcal{S}}$ is proportional to the variance in internal energy, i.e. $\tilde{C}_{\mathcal{S}} \simeq \beta^2 \Delta U_{\mathcal{S}}^2$ [2], with $\Delta U_{\mathcal{S}}^2 = \mathrm{Var}[\hat{\pi}_{\mathcal{S}}^0, \hat{H}_{\mathcal{S}}]$. This standard result is a form of fluctuation-dissipation relation; the rate of change of the system's energy with temperature is directly linked to the resulting energy fluctuations at equilibrium. Notably since the variance of a random variable is strictly non-negative, this implies $\tilde{C}_{\mathcal{S}} \geq 0$. This result tells us a basic fact about systems in thermal equilibrium, namely that heating up the system will increase its energy.

On the other hand, for finite coupling γ there are open quantum systems that may exhibit a negative specific heat at low temperatures, meaning that the fluctuation-dissipation relation cannot generally hold in this regime. Examples of such systems are the one-dimensional isotropic XY spin chain [90], a damped free particle [40], and a qubit coupled to a single bath resonator [89]. While perhaps unintuitive, the fact that $\tilde{C}_{\mathcal{S}} < 0$ is permissible implies that there are regions of temperature for which the system's energy actually increases as it is cooled down. The reason that one may obtain a negative heat capacity is because coupling the system to the reservoir may actually decrease the overall heat capacity due to the presence of non-vanishing interactions [40]. Analogously to (3.42), the heat capacity also satisfies the following additive relation:

$$\tilde{C}_{\mathcal{S}} = C_{\mathcal{S}\cup\mathcal{R}} - C_{\mathcal{R}}, \tag{3.47}$$

which clearly shows that $\tilde{C}_{\mathcal{S}}$ can be negative when the system is immersed in the reservoir. It is worth remarking that while a negative heat capacity may imply that the system is thermodynamically unstable, the fact that the total heat capacity of $\mathcal{S} \cup \mathcal{R}$ remains positive ensures that this is not the case [90].

While the standard fluctuation-dissipation relation is not generally fulfilled for finite γ, one can still derive a relation between the heat capacity and fluctuations in energy [15]. It is in this context that one uncovers a connection between thermodynamics and the average Wigner-Yanase-Dyson skew information (3.15), which is evidenced by the following theorem:

Theorem 3.4.1. *Consider the heat capacity* (3.135) *at arbitrary coupling strengths and denote* $\Delta\tilde{U}_{\mathcal{S}}^2 = \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*]$ *as the variance in the internal energy operator. Then*

$$\tilde{C}_{\mathcal{S}} = \beta^2 \Delta\tilde{U}_{\mathcal{S}}^2 - \beta^2 Q[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] + \left\langle \partial_T \hat{E}_{\mathcal{S}}^* \right\rangle^{eq}, \tag{3.48}$$

where $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*]$ *is the skew information defined in* (3.15).

Proof. Denote the operator $\delta\hat{E}_{\mathcal{S}}^* := \hat{E}_{\mathcal{S}}^* - \langle \hat{E}_{\mathcal{S}}^* \rangle$ as the deviation in internal energy, dropping the temperature dependence for now. Using (3.15) we now evaluate the average WYD skew information of the internal energy:

$$\begin{aligned}
Q[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] &= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - K[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - \int_0^1 da\ \mathrm{tr}[\hat{\pi}_{\mathcal{S}}^{1-a} \delta\hat{E}_{\mathcal{S}}^*\ \hat{\pi}_{\mathcal{S}}^{a}\ \delta\hat{E}_{\mathcal{S}}^*], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - \int_0^1 da\ \mathrm{tr}[e^{-(1-a)(\beta\hat{H}_{\mathcal{S}} + \ln \mathcal{Z}_{\mathcal{S}}^*)} \delta\hat{E}_{\mathcal{S}}^* e^{-a(\beta\hat{H}_{\mathcal{S}}^* + \ln Z_{\mathcal{S}}^*)} \delta\hat{E}_{\mathcal{S}}^*], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] + \mathrm{tr}[\delta\hat{E}_{\mathcal{S}}^*\ \partial_\beta \hat{\pi}_{\mathcal{S}}], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] + \mathrm{tr}[\hat{E}_{\mathcal{S}}^*\ \partial_\beta \hat{\pi}_{\mathcal{S}}], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - T^2 \mathrm{tr}[\hat{E}_{\mathcal{S}}^*\ \partial_T \hat{\pi}_{\mathcal{S}}], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - T^2 \partial_T \mathrm{tr}[\hat{E}_{\mathcal{S}}^*\ \hat{\pi}_{\mathcal{S}}] + T^2 \mathrm{tr}[\partial_T \hat{E}_{\mathcal{S}}^*\ \hat{\pi}_{\mathcal{S}}], \\
&= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - T^2 \tilde{C}_{\mathcal{S}} + T^2 \left\langle \partial_T \hat{E}_{\mathcal{S}}^* \right\rangle^{eq}, \qquad (3.49)
\end{aligned}$$

where we used the relation $\delta\hat{E}_{\mathcal{S}}^* = \partial_\beta(\beta\hat{H}_{\mathcal{S}}^* + \ln \mathcal{Z}_{\mathcal{S}}^*)$ and Lemma 3.4.1 in the the fourth line, and the fact that the operator $\partial_\beta \hat{\pi}_{\mathcal{S}}$ is traceless in the fifth line. Rearranging both sides completes the derivation. ■

Theorem 3.4.1 represents the most general equilibrium fluctuation-dissipation relation for systems coupled arbitrarily strong to a thermal reservoir. It is only in the limit $\gamma^2 \ll 1$ that one recovers the usual relation $\tilde{C}_{\mathcal{S}} = \beta^2 \tilde{U}_{\mathcal{S}}^2 \geq 0$. We see now that in the fully quantum regime, the heat capacity is influenced by additional *non-classical* fluctuations in energy, as quantified by the skew information (3.15). We have thus obtained a direct relationship between thermodynamics and measures of quantum uncertainty. Additionally, we find in Theorem 3.4.1 that the heat capacity is also influenced by an additional dissipation term $\left\langle \partial_T \hat{E}_{\mathcal{S}}^* \right\rangle^{eq}$ which stems from the temperature dependence of the internal energy operator. In regards to the possible negativity of $\tilde{C}_{\mathcal{S}}$, we can now identify the necessary and sufficient condition:

$$\tilde{C}_{\mathcal{S}} \leq 0 \iff Q[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*] - \tilde{U}_{\mathcal{S}}^2 \geq T^2 \left\langle \partial_T \hat{E}_{\mathcal{S}}^* \right\rangle^{eq}, \tag{3.50}$$

We note that since $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] \leq \tilde{U}^2_{\mathcal{S}}$, this implies that the dissipation term must be negative. This inequality also helps to explain why negative heat capacities can emerge at low temperatures. In this regime one expects the quantum coherences to dominate, in which case $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]$ will be very close to $\tilde{U}^2_{\mathcal{S}}$. It then follows that a small amount of negative dissipation $\left\langle \partial_T \hat{E}^*_{\mathcal{S}} \right\rangle^{eq}$ will result in a negative heat capacity. Clearly this is a consequence of the fact that the energy operator depends on temperature.

To illustrate the presence of quantum energy coherences, let us consider the QBM model (3.35). Recall that the HMF is given by $\hbar\omega_T \left(\hat{n}_T + \frac{1}{2}\right)$. After differentiating with respect to β, The internal energy operator here is given by

$$\hat{E}^*_{\mathcal{S}} = \alpha_T \hat{H}^*_{\mathcal{S}} - g_T \frac{\hat{a}^2_T + (\hat{a}^\dagger_T)^2}{2}, \tag{3.51}$$

where $\alpha_T = 1 - \frac{\omega'_T}{\omega_T} T$ and $g_T = \hbar\omega_T T \frac{A'_T}{A_T}$. Clearly it can be seen that $[\hat{H}^*_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] \neq 0$. To proceed one needs to obtain the elements $E_{nm} = |\langle n| \delta\hat{E}^* |m\rangle|^2$, where $\delta\hat{E}^* = \hat{E}^*_{\mathcal{S}} - \langle \hat{E}^*_{\mathcal{S}} \rangle$. Firstly one finds the following:

$$\begin{aligned}
\langle n| \delta\hat{E}^* |m\rangle &= \langle n| \hat{E}^*_{\mathcal{S}} |m\rangle - \langle \hat{E}^*_{\mathcal{S}} \rangle \delta_{n,m}, \\
&= \alpha_T \langle n| \hat{H}^*_{\mathcal{S}} |m\rangle - g_T \langle n| \frac{\hat{a}^2_T + (\hat{a}^\dagger_T)^2}{2} |m\rangle - \langle \hat{E}^*_{\mathcal{S}} \rangle \delta_{n,m}, \\
&= \alpha_T (\epsilon_n - \langle \hat{H}^*_{\mathcal{S}} \rangle) \delta_{n,m} - g_T \left(\frac{\sqrt{m}\sqrt{m-1}}{2} \delta_{n,m-2} + \frac{\sqrt{n}\sqrt{n-1}}{2} \delta_{m,n-2} \right),
\end{aligned} \tag{3.52}$$

where $\delta_{n,m}$ represents the Kronecker-Delta function. Squaring both sides yields

$$E_{nm} = \alpha^2_T (\epsilon_n - \langle \hat{H}^*_{\mathcal{S}} \rangle)^2 \delta_{n,m} + \frac{g^2_T}{4} \left((n+2)(n+1)\delta_{n+2,m} + (m+2)(m+1)\delta_{m+2,n} \right), \tag{3.53}$$

Finally, the average skew information can be obtained using Lemma 3.2.1:

$$\begin{aligned}
Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] &= \sum_{n<m} \left(p_n + p_m - \frac{2(p_n - p_m)}{\ln p_n - \ln p_m} \right) E_{nm}, \\
&= \frac{g^2_T}{4} \sum_{n=0}^{\infty} \left(p_n + p_{n+2} - \frac{2(p_n - p_{n+2})}{\ln p_n - \ln p_{n+2}} \right) (n+2)(n+1), \\
&= g^2_T \sinh\left(\frac{\beta\omega_T}{2} \right) \frac{e^{-\frac{\beta\omega_T}{2}}}{(1 - e^{-\beta\omega_T})^3} \left(1 + e^{-2\beta\omega_T} - \frac{T}{\omega_T} (1 - e^{-2\beta\omega_T}) \right).
\end{aligned} \tag{3.54}$$

Figure 3.2 shows the square root of this average skew information in units of $\hbar\omega$ as a function of temperature for different coupling strengths. As expected we see that the

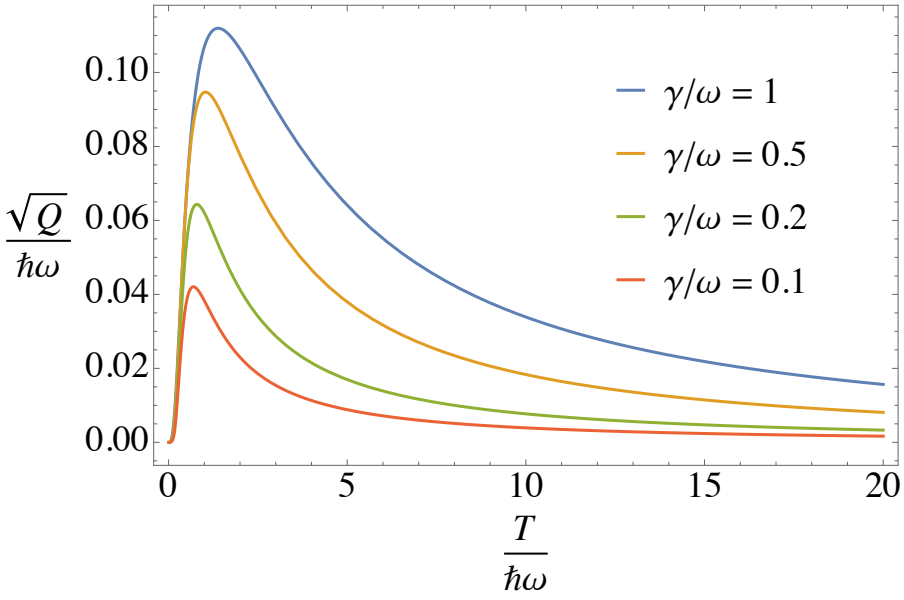


Figure 3.2: **Skew information for the damped oscillator.** Plot of quantum energetic fluctuations $\sqrt{Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]}/\hbar\omega$ for the damped oscillator as a function of $T/\hbar\omega$ for different coupling strengths γ. Here $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]$ is the average Wigner-Yanase-Dyson skew information for the effective energy operator $\hat{E}^*_{\mathcal{S}}$. These fluctuations are present when the state of the oscillator $\hat{\pi}_{\mathcal{S}}$ is not diagonal in the basis of $\hat{E}^*_{\mathcal{S}}$ due to the non-negligible interaction between the system and reservoir. The plot shows that increasing the coupling γ leads to an increase in the skew information. The quantum fluctuations are most pronounced at low temperatures where the thermal energies become comparable to the oscillator spacing, $T \simeq \hbar\omega$. As expected, the skew information decreases to zero in both the high temperature and weak coupling limits.

quantum fluctuations in energy vanish in the high temperature limit, while fluctuations grow with increased coupling strengths due to increased non-commutativity between $\hat{E}^*_{\mathcal{S}}$ and the state $\hat{\pi}_{\mathcal{S}}$ of the oscillator. Interestingly we see that $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]$ decays exponentially to zero in the low temperature limit, implying that the state of the oscillator commutes with the internal energy operator in this regime. Whether this is a general feature or specific to the example here remains an open question.

The quantum fluctuation-dissipation relation demonstrates that quantum coherences effect the thermodynamic properties of strongly-coupled systems. In the subsequent sections I will show that these coherences influence the fluctuations in the observable temperature of $\mathcal{S}$.

3.5 Quantum Estimation Theory

In this section I will outline the general parameter estimation problem in quantum mechanics. This will then be used to analyse the effects of quantum coherence on the estimation of temperature in strongly-coupled systems. Consider a quantum state $\hat{\rho}_\theta \in \mathcal{S}(\mathcal{H})$ that depends smoothly on some unknown parameter $\theta \in \mathbb{R}$. In general, such a parameter will be a non-linear function of the density operator, and so its value cannot be associated to a hermitian observable. Instead, in order to acquire information about θ one needs to perform a POVM measurement such as $\hat{M}_\zeta$, with $\int d\zeta \hat{M}_\zeta = \hat{\mathbb{I}}$. By Born's rule, a given outcome ζ from the data set conditioned on the value θ is given by $p(\zeta|\theta) = \mathrm{tr}[\hat{M}_\zeta \hat{\rho}_\theta]$. Within a given measurement setup, one may associate an estimator function $\breve{\theta}(\zeta)$ that provides a mapping from the set of outcomes ζ to some estimated value of θ. The accuracy of such an estimator is quantified by the *mean-squared-error* (MSE), which measures the difference between the estimate and the true parameter value:

$$\Delta\theta^2 := \int d\zeta \; p(\zeta|\theta)[\breve{\theta}(\zeta) - \theta]^2. \tag{3.55}$$

A particularly important choice of estimator is the *maximum-likelihood-estimator* (MLE), which is given by

$$\breve{\theta}_{\mathrm{MLE}}(\zeta) := \arg\max_\theta p(\zeta|\theta), \tag{3.56}$$

The MLE can be interpreted as the function that, for a given instance of outcomes ζ, outputs the value of parameter for which this data sample is most probable.

Let us further restrict our attention to estimators that are *unbiased*. This means that on average the estimator outputs the correct value, so that

$$\langle\breve{\theta}\rangle := \int d\zeta \; p(\zeta|\theta)\breve{\theta}(\zeta) = \theta. \tag{3.57}$$

In this case the MSE reduces to the variance of the chosen estimator:

$$\Delta\theta^2 = \int d\zeta \; p(\zeta|\theta)[\breve{\theta}(\zeta) - \langle\breve{\theta}\rangle]^2 = \mathrm{Var}[\breve{\theta}]. \tag{3.58}$$

It is important to note that in general, a given statistical model may not have a global estimator that is unbiased. For example, for a finite data set the MLE (3.56) can not be guaranteed to be unbiased. However, in the asymptotic limit in which the data set grows to infinity it can be shown that (3.56) indeed becomes unbiased, and we will focus

attention on this asymptotic regime. It follows from condition (3.57) and (3.58) that the MSE is bounded according to the Cramér-Rao inequality [46]:

$$\Delta\theta^2 \geq \frac{1}{\mathcal{F}(\theta)}, \tag{3.59}$$

where

$$\mathcal{F}(\theta) = \int d\zeta \; p(\zeta|\theta)\left(\frac{\partial \ln p(\zeta|\theta)}{\partial\theta}\right)^2, \tag{3.60}$$

is known as the *classical Fisher information*. This function essentially measures the sensitivity of the statistical model to changes in the underlying parameter θ.

So far we have fixed the POVM and the resulting distribution $p(\zeta|\theta)$. However, in quantum mechanics we have the additional freedom to choose our measurement basis, and one can seek to maximise the classical Fisher information in (3.60) over the set of POVMs. To do this we now introduce a hermitian operator known as the *symmetric-logarithmic-derivative* (SLD) [111], denoted $\hat{L}_\theta$, which is defined according to the following operator equation:

$$\frac{\partial}{\partial\theta}\hat{\rho}_\theta := \frac{1}{2}\{\hat{\rho}_\theta, \hat{L}_\theta\}, \tag{3.61}$$

Using this one may rewrite (3.60)

$$\mathcal{F}(\theta) = \int d\zeta \; \frac{\mathrm{Re}\big(\mathrm{tr}[\hat{M}_\zeta\hat{\rho}_\theta\hat{L}_\theta]\big)^2}{\mathrm{tr}[\hat{M}_\zeta\hat{\rho}_\theta]}, \tag{3.62}$$

Invoking the Cauchy-Schwartz inequality $|\mathrm{tr}[\hat{A}^\dagger\hat{B}]|^2 \leq \mathrm{tr}[\hat{A}^\dagger\hat{A}]\mathrm{tr}[\hat{B}^\dagger\hat{B}]$, one finds the following chain of inequalities [111, 112]:

$$\begin{aligned}
\mathcal{F}(\theta) &\leq \int d\zeta \; \left|\frac{\mathrm{tr}[\hat{M}_\zeta\hat{\rho}_\theta\hat{L}_\theta]}{(\mathrm{tr}[\hat{M}_\zeta\hat{\rho}_\theta])^{1/2}}\right|^2, \\
&= \int d\zeta \; \left|\mathrm{tr}\left[\frac{\hat{\rho}_\theta^{1/2}\hat{M}_\zeta^{1/2}}{(\mathrm{tr}[\hat{M}_\zeta\hat{\rho}_\theta])^{1/2}}.\hat{M}_\zeta^{1/2}\hat{L}_\theta\hat{\rho}_\theta^{1/2}\right]\right|^2, \\
&\leq \int d\zeta \; \mathrm{tr}[\hat{M}_\zeta\hat{L}_\theta\hat{\rho}_\theta\hat{L}_\theta], \\
&= \mathrm{tr}[\hat{L}_\theta^2 \; \hat{\rho}_\theta]. \qquad (3.63)
\end{aligned}$$

The quantity appearing in the final line is known as the *quantum Fisher information* (QFI) [113]:

$$\tilde{\mathcal{F}}(\theta) := \mathrm{tr}[\hat{L}_\theta^2 \; \hat{\rho}_\theta]. \tag{3.64}$$

Importantly, the QFI is a functional of the input state $\hat{\rho}_\theta$ and sets a bound on the minimum MSE independent of the measurement setup. If we also consider repeating the measurement process ν times, the additivity of the QFI and (3.63) implies the quantum Cramér-Rao inequality [111–113]:

$$\Delta\theta^2 \geq \frac{1}{\nu\tilde{\mathcal{F}}(\theta)}. \tag{3.65}$$

This bound is tighter than (3.59) due to the additional optimisation over the set of quantum measurements.

For a finite number of measurements ν, the bound (3.65) is not tight. However, in the asymptotic limit $\nu \to \infty$ one can indeed saturate the quantum Cramér-Rao inequality. First note that the chain of inequalities (3.63) become tight under the conditions

$$\mathrm{Im}\big(\mathrm{tr}[\hat{M}_\zeta \hat{\rho}_\theta \hat{L}_\theta]\big) = 0, \tag{3.66}$$

$$\frac{\hat{M}_\zeta^{1/2}\hat{\rho}_\theta^{1/2}}{\mathrm{tr}[\hat{M}_\zeta\hat{\rho}_\theta]} = \frac{\hat{M}_\zeta^{1/2}\hat{L}_\theta\hat{\rho}_\theta^{1/2}}{\mathrm{tr}[\hat{M}_\zeta\hat{L}_\theta\hat{\rho}_\theta]}. \tag{3.67}$$

It follows that these conditions are satisfied by choosing the measurement $\hat{M}_\zeta = \hat{P}_\zeta(\theta)$, with $\{\hat{P}_\zeta(\theta)\}$ the set of projectors onto the eigenstates of the symmetric-logarithmic-derivatives $\hat{L}_\theta$. This condition fixes the measurement basis, but does not determine the estimator.

As already pointed out, while measuring in the basis of $\hat{L}_\theta$ gives the optimal quantum measurement for single parameter estimation, in general there may not exist an unbiased estimator saturating (3.65) for finite repetitions ν. Furthermore, since the SLD will generally depend on θ, performing this measurement requires knowledge of the parameter itself. This contradicts the premise of the problem, given that θ is assumed unknown. Luckily, one may in fact circumvent these issues using an adaptive measurement scheme [114]. Firstly, an arbitrary guess is taken with estimate $\breve{\theta}_0 \in \mathbb{R}$ and the optimal measurement commuting with $\hat{L}_{\breve{\theta}_0}$ is performed. If outcome ζ_1 is obtained, then one applies the MLE (3.56) to likelihood function $p_{\breve{\theta}_0}(\zeta_1|\theta) = \mathrm{tr}[\hat{P}_{\zeta_1}(\breve{\theta}_0)\hat{\rho}_\theta]$, and arrives at the next guess $\breve{\theta}_1$. For the n'th ($\nu \geq 2$) step, a measurement projecting onto $\hat{L}_{\breve{\theta}_{\nu-1}}$ is performed with $\breve{\theta}_{\nu-1}$ is the MLE obtained at the previous stage. To obtain the ν'th MLE we choose the estimator

$$\breve{\theta}_\nu := \arg\max_\theta \prod_{i=1}^{\nu} p_{\breve{\theta}_{i-1}}(\zeta_i|\theta), \tag{3.68}$$

with ζ_i the data obtained at the i'th stage. Within this adaptive estimation scheme, it can be rigorously proven that this sequence of MLEs will converge to the true parameter in the sense that

$$\lim_{\nu\to\infty} \langle \breve{\theta}_\nu \rangle = \theta, \tag{3.69}$$

where the average is taken over the total data sequence [114]. Thus in the asymptotic limit the estimator becomes unbiased, and one may apply the quantum Cramér-Rao inequality (3.65). Furthermore, it follows that the bound is tight [96]. Denoting the MSE for this adaptive scheme by $\Delta\theta^2_{opt}$, we have

$$\lim_{\nu\to\infty} \nu \Delta\theta^2_{opt} = \frac{1}{\tilde{\mathcal{F}}(\theta)}. \tag{3.70}$$

Most importantly, this measurement scheme required no prior knowledge of the true parameter θ. Intuitively the convergence of the sequence can be shortened when the initial guess $\breve{\theta}_0$ is close to θ [115].

From this analysis, we can conclude that the QFI sets the ultimate achievable bound on asymptotic parameter estimation. This fundamental quantity plays a vital role in determining precision bounds in quantum phase estimation [79], estimating dissipative parameters in open quantum systems [116], anisotropy in spin-chains [117] and countless other situations. Within this chapter it will be shown that the QFI is intimately related to thermodynamics and limits the precision for estimates of temperature of quantum systems.

To gain further intuition for what the QFI is, it can be shown that it satisfies the following identity:

$$\tilde{\mathcal{F}}(\theta) = -2 \lim_{\delta\theta\to 0} \frac{\mathbb{F}(\hat{\rho}_{\theta+\delta\theta}, \hat{\rho}_\theta)}{\delta\theta^2}, \tag{3.71}$$

where

$$\mathbb{F}(\hat{\rho}, \hat{\sigma}) := \mathrm{tr}\left[\sqrt{\sqrt{\hat{\rho}}\hat{\sigma}\sqrt{\hat{\rho}}} \right], \tag{3.72}$$

is the *Ulhmann fidelity* [118]. The fidelity acts as a distinguishability measure on $\mathcal{S}(\mathcal{H})$, since it satisfies $0 \leq \mathbb{F}(\hat{\rho}, \hat{\sigma}) = \mathbb{F}(\hat{\sigma}, \hat{\rho}) \leq 1$ with $\mathbb{F}(\hat{\rho}, \hat{\sigma}) = 0$ if and only if $\hat{\rho} = \hat{\sigma}$. Thus we see that the QFI gauges the responsiveness of the quantum state to infinitesimal perturbations in parameter θ. To compute the QFI one needs to first find the SLD's in (3.61). This anti-commutator equation is essentially a particular instance of the

operator Lyapanov equation [119], and since $\hat{\rho}_\theta$ is a positive bounded operator there exists a unique solution given by

$$\hat{L}_\theta := 2\int_0^\infty d\lambda\; e^{-\lambda\hat{\rho}_\theta}\left(\frac{\partial}{\partial\theta}\hat{\rho}_\theta\right)\; e^{-\lambda\hat{\rho}_\theta}, \tag{3.73}$$

By substituting this into (3.64) one obtains the expression [120]

$$\tilde{\mathcal{F}}(\theta) = 2\int_0^\infty d\lambda\; \mathrm{tr}\left[\left(e^{-\lambda\hat{\rho}_\theta}\frac{\partial}{\partial\theta}\hat{\rho}_\theta\right)^2\right], \tag{3.74}$$

which is a closed expression for the QFI valid for states of arbitrary rank.

Let us now consider a simple example, where the parameter to be estimated is time t for a unitary evolution

$$\hat{\rho}_t = e^{it\hat{H}/\hbar}\hat{\rho}_0 e^{-it\hat{H}/\hbar}, \tag{3.75}$$

with $\hat{\rho}_0 = |\psi\rangle\langle\psi|$ some pure state and $\hat{H}$ a time-independent Hamiltonian. Using $\hat{\rho}_t = \hat{\rho}_t^2$ we find $\partial_t\hat{\rho}_t = \{\hat{\rho}_t, \partial_t\hat{\rho}_t\}$, and the uniqueness of the Lyapanov equation implies $\hat{L}_t = 2\partial_t\hat{\rho}_t$ for a pure state. This means that the SLD becomes

$$\hat{L}_t = \frac{2i}{\hbar}[\hat{\rho}_0, \hat{H}]. \tag{3.76}$$

Since the SLD is independent of t, this means that the optimal estimator is given by the MLE (3.56) with a projective measurement onto the eigenstates of the commutator $[\hat{\rho}_0, \hat{H}]$. The QFI (3.64) then reduces to the variance in the initial energy:

$$\tilde{\mathcal{F}} = \frac{4}{\hbar^2}\Delta E^2, \tag{3.77}$$

with $\Delta E^2 = \mathrm{Var}[\hat{\rho}_0, \hat{H}]$. Finally, from the quantum Cramér-Rao inequality (3.65) we get

$$\Delta E\Delta t \geq \frac{\hbar}{2}, \tag{3.78}$$

with Δt the uncertainty in the estimated time for a single measurement. The inequality (3.78) can be interpreted as a type of *energy-time uncertainty relation* [112], and reveals that increasing the initial energy spread of a pure state reduces the uncertainty in the observable time of its subsequent evolution. Note that this type of quantum trade-off is fundamentally different from the familiar Heisenberg uncertainty relation, due to the fact that time is a deterministic parameter rather than a quantum observable. Here the uncertainty Δt stems from the limitations of a particular estimation strategy adopted by the observer, rather than an intrinsic property of the quantum state such as ΔE.

3.6 Energy-temperature uncertainty relation

The formalism presented in the previous section describes estimation of arbitrary parameters in quantum systems. However, our focus will be on estimating temperatures of quantum systems. In standard thermometry, a thermometer is placed in contact with the target sample and is left to reach thermal equilibrium at temperature T. The temperature of the sample can be inferred from the state of the probe through suitable processing. In a quantum scenario the same procedure can be applied, where the probe is now taken to be a microscopic quantum system such as a single ion or impurity in a nano-crystal. This setup is important when the sample is a quantum many-body system such as a Bose-Einstein condensate or spin-chain, in which case a non-invasive means of estimating temperature is desirable in order to avoid disturbing the sample.

To analyse the performance of a particular thermometry scheme, we will consider the system $\mathcal{S}$ as our probe and $\mathcal{R}$ as the sample. After sufficient time the interaction between $\mathcal{S}$ and $\mathcal{R}$ will cause the probe to approach the state (3.25). The optimal estimate of temperature from a local measurement of $\mathcal{S}$ can be determined via the quantum Fisher information (3.64) with respect to parameter T. As is customary in standard thermodynamics, we first assume weak-coupling $\gamma^2 \ll 1$. In this case the SLD (3.73) is found to be

$$\hat{L}_{\mathcal{S}}(T) = \frac{\hat{H}_{\mathcal{S}} - U_{\mathcal{S}}}{T^2}, \tag{3.79}$$

which immediately implies that the optimal quantum estimator is given by a projective measurement onto the energy basis of the bare Hamiltonian $\hat{H}_{\mathcal{S}}$. The QFI is then found to be proportional to the energy variance in the system probe:

$$\tilde{\mathcal{F}}(T) = \frac{\Delta U_{\mathcal{S}}^2}{T^4}. \tag{3.80}$$

Note that the QFI for T and inverse β are simply related according to

$$\tilde{\mathcal{F}}(T) = \frac{\tilde{\mathcal{F}}(\beta)}{T^4}, \tag{3.81}$$

Let us now denote the fluctuations in temperature and inverse temperature by $\Delta T_{\mathcal{S}}$ and $\Delta\beta_{\mathcal{S}}$ respectively for a particular estimation scheme with a single measurement. Using the Cramér-Rao inequality (3.65) leads us to the famous *thermodynamic uncertainty principle*:

$$\Delta\beta_{\mathcal{S}}\Delta U_{\mathcal{S}} \geq 1. \tag{3.82}$$

This relation implies that increasing the energy spread of the probe reduces the minimal fluctuations in observable temperature. Alternatively, one may utilise the fluctuation-dissipation relation, $C_{\mathcal{S}} = \beta^2 \Delta U_{\mathcal{S}}^2$ to rewrite (3.82) in terms of the system heat capacity:

$$\left(\frac{T}{\Delta T_{\mathcal{S}}}\right)^2 \leq C_{\mathcal{S}}. \tag{3.83}$$

Thus the most accurate probes are those with a large heat capacity. Intuitively this makes sense, as this relation implies that the best thermometers have energies that are highly-sensitive to small changes in temperature. For simple probe systems with $U_{\mathcal{S}} \propto T$ such as a harmonic oscillator, the optimal estimator $\breve{T}$ is given by the energy eigenvalues themselves. The optimal probe that gives the best signal-to-noise estimate of T is found from (3.83) by maximising the heat capacity. For a finite-dimensional quantum probe this optimisation has been achieved, and the optimal quantum thermometer becomes an effective two-level system with a maximally degenerate excited state [70].

The bounds (3.82) and (3.83) place fundamental restrictions on the precision of temperature measurements of both quantum and classical systems within the weak-coupling approximation. However, it is clear from the above derivation that these relations will no longer hold when the coupling term γ cannot be neglected. In this regime we know that the probe state will deviate from a local Gibbs state, and is more generally given by the reduced thermal state (3.25). It is unclear how the temperature fluctuations relate to the energy of a general strongly-coupled probe. By bringing the results of the previous sections together, this question is answered by the following theorem:

Theorem 3.6.1. *Let the state of the probe $\mathcal{S}$ be given by* (3.25) *with γ arbitrarily strong, and $\Delta\beta_{\mathcal{S}}$ the uncertainty in β obtained from a local measurement of $\mathcal{S}$. Then*

$$\Delta\beta_{\mathcal{S}} \geq \frac{1}{\sqrt{\tilde{U}_{\mathcal{S}}^2 - Q[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*]}} \geq \frac{1}{\tilde{U}_{\mathcal{S}}}, \tag{3.84}$$

with $\tilde{U}_{\mathcal{S}}^2 = \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}_{\mathcal{S}}^*]$ the variance in the internal energy operator $\hat{E}_{\mathcal{S}}^*$.

Proof. We begin by considering the reduced thermal state $\hat{\pi}_{\mathcal{S}}(\beta) = e^{-\beta \hat{H}_{\mathcal{S}}^*}/\mathcal{Z}_{\mathcal{S}}^*$. Suppressing the dependence on β for now, let us denote the spectral decomposition by $\hat{\pi}_{\mathcal{S}} = \sum_n p_n |\psi_n\rangle \langle\psi_n|$ where the eigenstates satisfy $\hat{H}_{\mathcal{S}}^* |\psi_n\rangle = \lambda_n |\psi_n\rangle$. We arrange the sum in decreasing order, so that $p_n \geq p_m$ if $n < m$. The QFI with respect to β is then given by (3.74). By expanding in the basis of $\hat{\pi}_{\mathcal{S}}$ and evaluating the integral in (3.74),

the QFI can be written as follows :

$$\tilde{\mathcal{F}}(\beta) = 2\sum_{n,m} \frac{|\langle\psi_n|\,\partial_\beta\hat{\pi}_{\mathcal{S}}\,|\psi_m\rangle\,|^2}{p_n+p_m}. \tag{3.85}$$

Given the exponential form of $\hat{\pi}_{\mathcal{S}}$, we use Lemma 3.4.1 to expand the derivative:

$$\partial_\beta\ \hat{\pi}_{\mathcal{S}} := \frac{-1}{\mathcal{Z}^*_{\mathcal{S}}}\int_0^1 da\ e^{-(1-a)\beta\hat{H}^*_{\mathcal{S}}}\ \delta\hat{E}^*_{\mathcal{S}}\ e^{-a\beta\hat{H}^*_{\mathcal{S}}}, \tag{3.86}$$

where $\delta\hat{E}^*_{\mathcal{S}} = \hat{E}^*_{\mathcal{S}} - \tilde{U}_{\mathcal{S}}$. Using this the QFI becomes

$$\begin{aligned}
\tilde{\mathcal{F}}(\beta) &= 2\sum_{n,m}\frac{|\langle\psi_n|\,\partial_\beta\hat{\pi}_{\mathcal{S}}\,|\psi_m\rangle\,|^2}{p_n+p_m},\\
&= \frac{2}{(\mathcal{Z}^*_{\mathcal{S}})^2}\sum_{n,m}\frac{1}{p_n+p_m}\left|\langle\psi_n|\int_0^1 da\ e^{-(1-a)\beta\hat{H}^*_{\mathcal{S}}}\ \delta\hat{E}^*_{\mathcal{S}}\ e^{-a\beta\hat{H}^*_{\mathcal{S}}}\,|\psi_m\rangle\right|^2,\\
&= \sum_n p_n\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_n\rangle\,\big|^2\\
&\qquad + \frac{4}{(\mathcal{Z}^*_{\mathcal{S}})^2}\sum_{n<m}\frac{1}{p_n+p_m}\big|\,\langle\psi_m|\,\hat{E}^*_{\mathcal{S}}\,|\psi_n\rangle\,\big|^2\left[\int_0^1 da\ e^{-\left(a\lambda_m+(1-a)\lambda_n\right)}\right]^2,\\
&= \sum_n p_n\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_n\rangle\,\big|^2 + 4\sum_{n<m}\frac{(p_n-p_m)^2}{p_n+p_m}\big|\,\langle\psi_n|\,\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2\frac{1}{(\ln p_n-\ln p_m)^2},
\end{aligned} \tag{3.87}$$

Let us now use the following expression for the variance:

$$\begin{aligned}
\mathrm{Var}[\hat{\pi}_{\mathcal{S}},\hat{E}^*_{\mathcal{S}}] &= \sum_{n,m}\frac{p_n+p_m}{2}\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2,\\
&= \sum_n p_n\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_n\rangle\,\big|^2 + \sum_{n\neq m}\frac{p_n+p_m}{2}\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2,\\
&= \sum_n p_n\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_n\rangle\,\big|^2 + \sum_{n<m}(p_n+p_m)\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2,
\end{aligned} \tag{3.88}$$

Comparing this with (3.87) we now add and subtract the sum $\sum_{n<m}(p_n+p_m)\big|\,\langle\psi_n|\,\delta\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2$ to the RHS of (3.87), obtaining

$$\begin{aligned}
\tilde{\mathcal{F}}(\beta) = \mathrm{Var}[\hat{\pi}_{\mathcal{S}},\hat{E}^*_{\mathcal{S}}] + \sum_{n<m}\Bigg[&\left(\frac{2(p_n-p_m)}{\ln(p_n/p_m)}\right)\left(\frac{2(p_n-p_m)}{(p_n+p_m)\ln(p_n/p_m)}\right)\\
&- (p_n+p_m)\Bigg]\big|\,\langle\psi_n|\,\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2,
\end{aligned} \tag{3.89}$$

We now turn to the average WYD skew information (3.15) of observable $\hat{E}^*_{\mathcal{S}}$, which is given by Lemma 3.2.1:

$$Q[\hat{\pi}_{\mathcal{S}},\hat{E}^*_{\mathcal{S}}] = \sum_{n<m}\left(p_n+p_m-\frac{2(p_n-p_m)}{\ln p_n-\ln p_m}\right)\big|\,\langle\psi_n|\,\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\,\big|^2. \tag{3.90}$$

We now bound the QFI following on from (3.122):

$$\begin{aligned}\tilde{\mathcal{F}}(\beta) &= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] + \sum_{n<m}\left[\left(\frac{2(p_n-p_m)}{\ln(p_n/p_m)}\right)\left(\frac{2(p_n-p_m)}{(p_n+p_m)\ln(p_n/p_m)}\right)\right. \\ &\qquad\qquad \left. - (p_n+p_m)\right]\left|\langle\psi_n|\,\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\right|^2, \\ &\leq \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] + \sum_{n<m}\left[\frac{2(p_n-p_m)}{\ln(p_n/p_m)} - (p_n+p_m)\right]\left|\langle\psi_n|\,\hat{E}^*_{\mathcal{S}}\,|\psi_m\rangle\right|^2, \\ &= \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] - Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}], \end{aligned} \tag{3.91}$$

where in the second line we used the fact that $(p_n - p_m)/\ln(p_n/p_m) \geq 0$ since $p_n \geq p_m$ for $n < m$, and the inequality

$$\frac{x-1}{x+1} \leq \ln\sqrt{x}; \qquad x \geq 1, \tag{3.92}$$

identifying $x = p_n/p_m \geq 1$. This allowed us to use

$$\left(\frac{2(p_n-p_m)}{(p_n+p_m)\ln(p_n/p_m)}\right) \leq 1, \tag{3.93}$$

for each term inside the sum. In the third line we used the expression (3.90) for the skew information. This completes the proof. ■

This theorem represents the strong-coupling generalisation of (3.82). It can be seen that the bound on the uncertainty in the inverse temperature is increased whenever quantum energy fluctuations are present for a fixed amount of total energy variance. These additional fluctuations are quantified by the non-negative skew information $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]$ defined in (3.15), and this contribution provides a tighter bound on $\Delta\beta_{\mathcal{S}}$ than (3.82). One recovers the usual uncertainty relation when $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]$ can be neglected, which is the case when the interaction commutes with the bare Hamiltonian of $\mathcal{S}$ and $\mathcal{R}$ or when the interaction is sufficiently weak. We note that $Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]$ vanishes for classical systems and Theorem 3.6.1 reduces to the original uncertainty relation (3.82), but with energy fluctuations quantified by $\hat{E}^*_{\mathcal{S}}$ instead of the bare Hamiltonian $\hat{H}_{\mathcal{S}}$.

With the strong-coupling energy-temperature uncertainty relation established, the next question that arises is how to relate these quantities to the heat capacity and generalise (3.83). This can be achieved using the fluctuation-dissipation relation in Theorem 3.4.1:

Theorem 3.6.2. *Let $\Delta T_{\mathcal{S}}$ denote the uncertainty in a temperature estimate obtained via the probe state* (3.25) *at arbitrary coupling. Then the signal-to-noise ratio is upper bounded as follows:*

$$\left(\frac{T}{\Delta T_{\mathcal{S}}}\right)^2 \leq \tilde{C}_{\mathcal{S}} - \left\langle \partial_T \hat{E}^*_{\mathcal{S}} \right\rangle^{eq}. \tag{3.94}$$

Proof. From Theorem 3.6.1 we have the bound

$$\tilde{\mathcal{F}}(\beta) \leq \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] - Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]. \tag{3.95}$$

Furthermore, changing variables $\beta \to T$ is achieved via $\tilde{\mathcal{F}}(\beta) = T^4 \tilde{\mathcal{F}}(T)$. Using the FDR in Theorem 3.4.1 then yields

$$\tilde{\mathcal{F}}(T) \leq \frac{\mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] - Q[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}]}{T^4}, \tag{3.96}$$

$$= \frac{\tilde{C}_{\mathcal{S}} - \left\langle \partial_T \hat{E}^*_{\mathcal{S}} \right\rangle^{eq}}{T^2}. \tag{3.97}$$

Finally, combination with the quantum Cramér-Rao inequality (3.65) completes the derivation. ■

This result demonstrates that the optimal signal-to-noise ratio for estimating the temperature of $\mathcal{S}$ is bounded by both the heat capacity of the strongly-coupled probe and the added dissipation term, which can be both positive or negative. This bound is independently tight in both the high temperature and weak-coupling limits. It should also be noted that Theorem 3.6.2 is valid in the classical limit, in which case it is always tight and can be saturated using the MLE applied to a projective measurement of the operator $\hat{E}^*_{\mathcal{S}}$.

Theorem 3.6.1 and Theorem 3.6.2 establish the link between temperature uncertainty, energy fluctuations and heat capacity within the general strong coupling limit. In this limit we now see that additional constraints on temperature estimation arise due to the non-commutativity between $\hat{E}^*_{\mathcal{S}}$ and $\hat{H}^*_{\mathcal{S}}$, and also the additional temperature dependence of the effective internal energy operator $\hat{E}^*_{\mathcal{S}}$.

To investigate the tightness of the bound in Theorem 3.6.2 let us return to the example of the damped harmonic oscillator (3.35). The elements of the associated internal energy operator (3.51) have already been computed in (3.53) as well as the skew information

contribution (3.54). We next need the total variance in energy, which is given by

$$\begin{aligned}
\mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{E}^*_{\mathcal{S}}] &= \sum_{n,m=0}^{\infty} p_n E_{nm}, \\
&= \alpha_T^2 \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] + \frac{g_T^2}{2} \sum_{n=0}^{\infty} p_n (n+1)(n+2), \\
&= \alpha_T^2 \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] + g_T^2 \sinh\left(\frac{\beta\omega_T}{2}\right) \sum_{n=0}^{\infty} (n+1)(n+2) e^{-\beta\omega_T(n+\frac{1}{2})} \\
&= \alpha_T^2 \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] + 2g_T^2 \sinh\left(\frac{\beta\omega_T}{2}\right) \left(\frac{e^{-\frac{\beta\omega_T}{2}}}{(1-e^{-\beta\omega_T})^3}\right), \\
&= \frac{\alpha_T^2 \omega_T^2}{4 \sinh^2\left(\frac{\beta\omega_T}{2}\right)} + 2g_T^2 \sinh\left(\frac{\beta\omega_T}{2}\right) \left(\frac{e^{-\frac{\beta\omega_T}{2}}}{(1-e^{-\beta\omega_T})^3}\right), \qquad (3.98)
\end{aligned}$$

where we used the series $\sum_{n=0}^{\infty}(n+1)(n+2)x^n = 2/(1-x)^3$ for $|x| < 1$ and that the variance of $\hat{H}^*_{\mathcal{S}}$ is

$$\mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] = \frac{\omega_T^2}{4 \sinh^2\left(\frac{\beta\omega_T}{2}\right)}. \qquad (3.99)$$

We now compute the QFI, using (3.87) and the fact that $p_n \pm p_{n+2} = p_n(1 \pm e^{-2\beta\omega_T})$:

$$\begin{aligned}
T^4 \tilde{\mathcal{F}}(T) &= \sum_{n=0}^{\infty} p_n E_{nn} + 4 \sum_{n<m} \frac{(p_n - p_m)^2}{(p_n + p_m) \ln^2(\frac{p_n}{p_m})} E_{nm}, \\
&= \alpha_T^2 \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] + \frac{g_T^2}{\beta^2\omega_T^2} \sum_{n=0}^{\infty} \frac{(p_n - p_{n+2})^2}{p_n + p_{n+2}} (n+2)(n+1), \\
&= \alpha_T^2 \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] + \frac{g_T^2 (1-e^{-2\beta\omega_T})^2}{\beta^2\omega_T^2 (1+e^{-2\beta\omega_T})} \sum_{n=0}^{\infty} p_n (n+2)(n+1), \\
&= \alpha_T^2 \mathrm{Var}[\hat{\pi}_{\mathcal{S}}, \hat{H}^*_{\mathcal{S}}] + \frac{g_T^2}{\beta^2\omega_T^2} \sinh\left(\frac{\beta\omega_T}{2}\right) \left(\frac{(1-e^{-2\beta\omega_T})^2}{(1+e^{-2\beta\omega_T})}\right) \left(\frac{e^{-\frac{\beta\omega_T}{2}}}{(1-e^{-\beta\omega_T})^3}\right), \\
&= \frac{\alpha_T^2 \omega_T^2}{4 \sinh^2\left(\frac{\beta\omega_T}{2}\right)} + \frac{g_T^2}{\beta^2\omega_T^2} \sinh\left(\frac{\beta\omega_T}{2}\right) \left(\frac{(1-e^{-2\beta\omega_T})^2}{(1+e^{-2\beta\omega_T})}\right) \left(\frac{e^{-\frac{\beta\omega_T}{2}}}{(1-e^{-\beta\omega_T})^3}\right), \qquad (3.100)
\end{aligned}$$

Figure 3.3 shows the optimal signal-to-noise ratio for estimating T determined by the Cramér-Rao bound (3.65), $(T/\Delta T_{\mathcal{S}})^2_{\mathrm{opt}} = T^2\tilde{\mathcal{F}}(T)$, as a function of temperature T and coupling strength γ. The bound given by Theorem 3.6.2 in terms of the heat capacity and the additional dissipation term is plotted and shows very good agreement with the optimum estimation scheme quantified by the quantum Fisher information. The

bound clearly becomes tight in the high-temperature limit ($T \to \infty$) independent of the coupling strength. Conversely the bound is also tight in the weak-coupling limit ($\gamma \to 0$) independent of the temperature. The optimum and the bound both converge exponentially to zero as $T \to 0$, albeit with different rates of decay. Outside of these limits the difference between the bound and $(T/\Delta T_{\mathcal{S}})^2_{\text{opt}}$ has a maximum, and at the temperature and coupling for which this maximum occurs the bound is roughly 30% greater than $(T/\Delta T_{\mathcal{S}})^2_{\text{opt}}$.

3.7 Quantum thermometry at strong coupling

The most immediate question that one might have following Theorem 3.6.1 and Theorem 3.6.2 is whether or not the finite interaction brings an enhancement to the precision of the temperature estimate. If this were true, then one could engineer a strong interaction between probe and sample in order to gain more information about the temperature, as depicted in Figure 3.4. Unfortunately, the expressions appearing in Theorem 3.6.1 and Theorem 3.6.2 are generally not analytically available outside of simple models such as the damped harmonic oscillator. To circumvent this issue we can treat the interaction perturbatively to obtain an analytic formula for the probe's optimal signal-to-noise ratio up to second order in coupling for a very wide class of many-body systems.

We first need to find the reduced state (3.25) up to second order in γ. To derive such an expression one needs the following Taylor expansion for the operator exponential function [121]:

Lemma 3.7.1. *Let $\hat{A}, \hat{B}$ be a pair of linear operator acting on $\mathcal{S}$. Then the function $\mathbb{R} \ni \gamma \mapsto e^{\hat{A}+\gamma\hat{B}}$ can be expanded as*

$$e^{\hat{A}+\gamma\hat{B}} = \sum_{n=0}^{\infty} \hat{A}_n(1)\gamma^n, \tag{3.101}$$

where $\hat{A}_0(s) = e^{s\hat{A}}$ and

$$\hat{A}_n(s) := \int_0^s dt_1 \int_0^{t_1} dt_2 \ldots \int_0^{t_{n-1}} dt_n \; e^{(s-t_1)\hat{A}}\hat{B}e^{(t_1-t_2)\hat{A}}\hat{B}\ldots\hat{B}e^{t_n\hat{A}}. \tag{3.102}$$

For a straightforward proof of Lemma 3.7.1, see Thm 3.10 in [122].

Let us now assume that the interaction Hamiltonian between a $\mathcal{S}$ and $\mathcal{R}$ can be expressed as a tensor product of system and reservoir terms:

$$\hat{V}_{\mathcal{S}\cup\mathcal{R}} = \hat{V}_{\mathcal{S}} \otimes \hat{V}_{\mathcal{R}}. \tag{3.103}$$

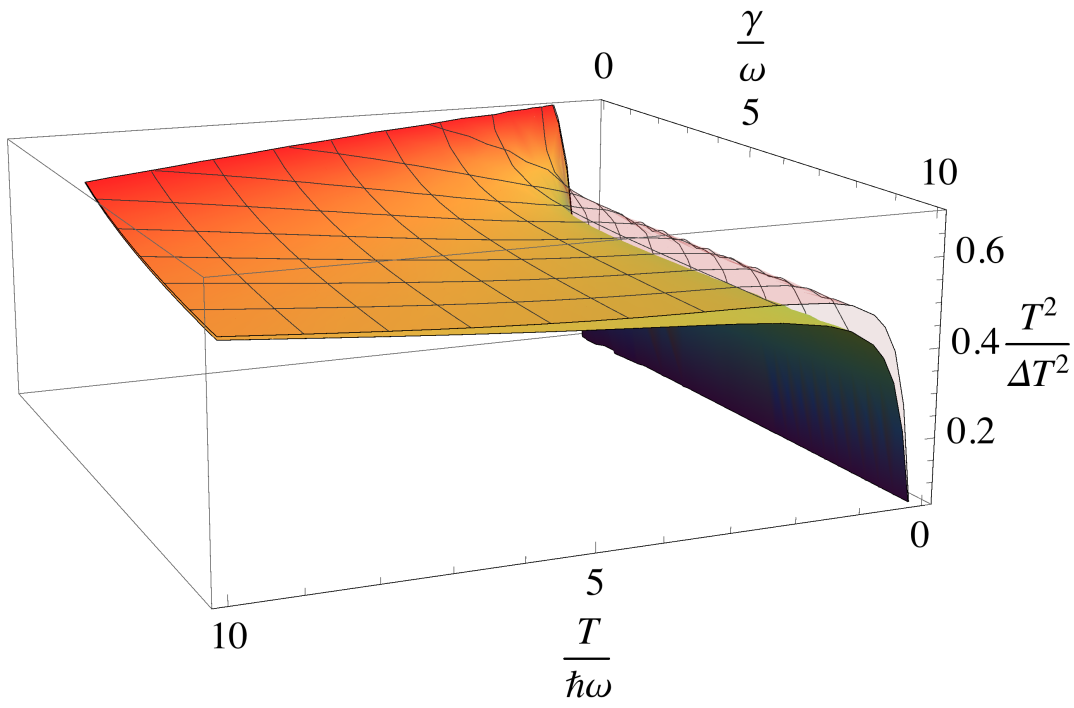


Figure 3.3: **Bound on temperature signal-to-noise ratio.** The coloured plot shows the optimal signal-to-noise ratio $(T/\Delta T_s)^2_{\text{opt}}$ of an unbiased temperature estimate for the damped oscillator, as a function of temperature T and coupling strength γ. This optimal measurement is determined by the quantum Fisher information, which places an asymptotically achievable lower bound on the temperature fluctuations ΔT_s through the Cramér-Rao inequality. The mesh plot shows the upper bound on $(T/\Delta T_s)^2_{\text{opt}}$ derived here from the generalised thermodynamic uncertainty relation in Theorem 3.6.2. This uncertainty relation links the temperature fluctuations to the heat capacity of the system at arbitrary coupling strengths. It can be seen that the upper bound becomes tight in both the high temperature and weak coupling limits.

While more general interactions can be considered, this form describes the vast majority of open systems encountered in physics [123]. Notably one may also assume that the $\mathcal{R}$

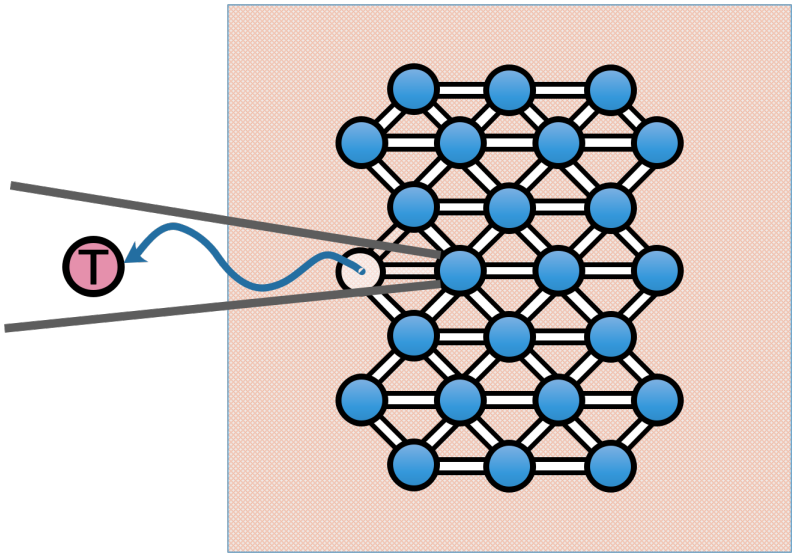


Figure 3.4: **Quantum thermometry.** The goal is to estimate the temperature of a many-body quantum system such as a Bose-Einstein condensate or spin chain. To avoid disturbing the sample one needs to attach a microscopic quantum probe to the sample. After equilibration, the probe encodes information about the temperature that can be processed via measurements of its state. If the interaction strength is arbitrarily large, the state of the probe will take the form (3.25), and the interaction will generally effect the accuracy of any temperature estimate. Depending on the form of the interaction and temperature, this additional dissipation may either be beneficial or detrimental to the probe's performance.

term averages to zero with respect to the reservoir degrees of freedom:

$$\langle \hat{V}_{\mathcal{R}} \rangle_{\mathcal{R}} = 0. \tag{3.104}$$

This standard assumption is trivial in the sense that one may always satisfy (3.104) by shifting the system Hamiltonian by this constant term [124]. Following these simple assumptions one can obtain the state (3.25) up to second order in coupling:

Theorem 3.7.1. *Let $\Phi_{\mathcal{S}}(t)$ and $\Phi_{\mathcal{R}}(t)$ denote the respective system and reservoir auto-correlation functions, with*

$$\Phi_\alpha(t) := \mathrm{tr}_\alpha[e^{it\hat{H}_\alpha}\hat{V}_\alpha e^{-it\hat{H}_\alpha}\hat{V}_\alpha\ \hat{\pi}^0_\alpha]; \quad \alpha = \mathcal{S}, \mathcal{R}. \tag{3.105}$$

Then the reduced state of $\mathcal{S}$ (3.25) is approximately given by

$$\hat{\pi}_{\mathcal{S}} = \hat{\pi}^0_{\mathcal{S}}\big(\hat{\mathbb{I}}_{\mathcal{S}} + \gamma^2 \hat{X}_{\mathcal{S}}\big) + \mathcal{O}(\gamma^3), \tag{3.106}$$

where

$$\hat{X}_{\mathcal{S}} = \int_0^{\beta} d\beta_1 \int_0^{\beta_1} d\beta_2 \; \Phi_{\mathcal{R}}(-i(\beta_1-\beta_2)) \times \left(e^{\beta_1 \hat{H}_{\mathcal{S}}} \hat{V}_{\mathcal{S}} e^{-(\beta_1-\beta_2))\hat{H}_{\mathcal{S}}} \hat{V}_{\mathcal{S}} e^{-\beta_2 \hat{H}_{\mathcal{S}}} - \Phi_{\mathcal{S}}(-i(\beta_1-\beta_2))\hat{\mathbb{I}}_{\mathcal{S}} \right), \tag{3.107}$$

Note that $tr_{\mathcal{S}}[\hat{X}_{\mathcal{S}}\hat{\pi}_{\mathcal{S}}] = 0$ *as required by normalisation.*

Proof. Applying Lemma 3.7.1 to $e^{-\beta \hat{H}_{\mathcal{S}\cup\mathcal{R}}}$ leads to:

$$\begin{aligned} e^{-\beta \hat{H}_{\mathcal{S}\cup\mathcal{R}}} =& e^{-\beta(\hat{H}_{\mathcal{S}}+\hat{H}_{\mathcal{R}})} \Bigg(1 - \gamma \int_0^{\beta} d\beta_1 \; e^{\beta_1(\hat{H}_{\mathcal{S}}+\hat{H}_{\mathcal{R}})} \hat{V}_{\mathcal{S}\cup\mathcal{R}} e^{-\beta_1(\hat{H}_{\mathcal{S}}+\hat{H}_{\mathcal{R}})} \\ &+ \gamma^2 \int_0^{\beta} d\beta_1 \int_0^{\beta_1} d\beta_2 \; e^{\beta_1(\hat{H}_{\mathcal{S}}+\hat{H}_{\mathcal{R}})} \hat{V}_{\mathcal{S}\cup\mathcal{R}} e^{-(\beta_1-\beta_2)(\hat{H}_{\mathcal{S}}+\hat{H}_{\mathcal{R}})} \hat{V}_{\mathcal{S}\cup\mathcal{R}} e^{-\beta_2(\hat{H}_{\mathcal{S}}+\hat{H}_{\mathcal{R}})} \Bigg) \\ &+ \mathcal{O}(\gamma^3). \end{aligned} \tag{3.108}$$

Taking the trace yields an approximation for the total partition function:

$$\begin{aligned} \mathcal{Z}_{\mathcal{S}\cup\mathcal{R}} \simeq& \mathcal{Z}_{\mathcal{S}}\mathcal{Z}_{\mathcal{R}} - \gamma\beta \mathrm{tr}_{\mathcal{S}}[e^{-\beta\hat{H}_{\mathcal{S}}}\hat{V}_{\mathcal{S}}]\mathrm{tr}_{\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{R}}}\hat{V}_{\mathcal{R}}] \\ &+ \gamma^2 \int_0^{\beta} d\beta_1 \int_0^{\beta_1} d\beta_2 \; \mathrm{tr}_{\mathcal{S}}[e^{-\beta\hat{H}_{\mathcal{S}}} e^{(\beta_1-\beta_2)\hat{H}_{\mathcal{S}}} \hat{V}_{\mathcal{S}} e^{-(\beta_1-\beta_2)\hat{H}_{\mathcal{S}}} \hat{V}_{\mathcal{S}}] \\ &\times \mathrm{tr}_{\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{R}}} e^{(\beta_1-\beta_2)\hat{H}_{\mathcal{R}}} \hat{V}_{\mathcal{R}} e^{-(\beta_1-\beta_2)\hat{H}_{\mathcal{R}}} \hat{V}_{\mathcal{R}}] \end{aligned} \tag{3.109}$$

Using the expansion $(a+b\gamma^2)^{-1} = a^{-1} - ba^{-2}\gamma^2 + \mathcal{O}(\gamma^3)$, and the fact that $\mathrm{tr}_{\mathcal{R}}[e^{-\beta\hat{H}_{\mathcal{R}}}\hat{V}_{\mathcal{R}}] = 0$ due to assumption (3.104), we find

$$\frac{1}{\mathcal{Z}_{\mathcal{S}\cup\mathcal{R}}} = \frac{1}{\mathcal{Z}_{\mathcal{S}}\mathcal{Z}_{\mathcal{R}}} \left(1 - \gamma^2 \int_0^{\beta} d\beta_1 \int_0^{\beta_1} d\beta_2 \; \Phi_{\mathcal{S}}(-i(\beta_1-\beta_2))\Phi_{\mathcal{R}}(-i(\beta_1-\beta_2)) \right) + \mathcal{O}(\gamma^3),$$

Multiplying the approximate $e^{-\beta\hat{H}_{\mathcal{S}\cup\mathcal{R}}}$ and $1/\mathcal{Z}_{\mathcal{S}\cup\mathcal{R}}$ together, collecting terms and taking the partial trace over the bath, we find that the first order correction vanishes. This completes the derivation. ■

For a finite but small coupling, the benefit of Theorem 3.7.1 is that one can analytically compute $\hat{\pi}_{\mathcal{S}}$ so long as the eigenstates of the bare system Hamiltonian are known along with the bath auto-correlation function. We now use Theorem 3.7.1 to derive the optimal temperature estimate for a coupled probe up to second order in γ.

Theorem 3.7.2. *Consider the temperature estimation scheme outlined in Section 3.5 applied to the coupled probe* (3.25)*, and denote the uncertainty after ν measurements by $\Delta_\nu T_{\mathcal{S}}$. We then denote the optimal signal-to-noise ratio in the asymptotic limit by*

$$\left(\frac{T}{\Delta T_{\mathcal{S}}}\right)^2_{opt} := \lim_{\nu\to\infty}\frac{1}{\nu}\left(\frac{T}{\Delta_\nu T_{\mathcal{S}}}\right)^2. \tag{3.110}$$

Then

$$\left(\frac{T}{\Delta T_{\mathcal{S}}}\right)^2_{opt} = C_{\mathcal{S}}(T) + \frac{\gamma^2}{T^2}\xi(T) + \mathcal{O}(\gamma^3), \tag{3.111}$$

where $C_{\mathcal{S}}(T)$ is the heat capacity of the uncoupled probe and

$$\xi(T) := \int_0^\beta du\ \left[(\beta-u)\Phi_{\mathcal{R}}(-iu)\frac{\partial^2}{\partial\beta^2}\Phi_{\mathcal{S}}(-iu) + 2\frac{\partial}{\partial\beta}\Phi_{\mathcal{S}}(-iu)\frac{\partial}{\partial\beta}(\beta-u)\Phi_{\mathcal{R}}(-iu)\right]. \tag{3.112}$$

Proof. Using Theorem 3.7.1 we have

$$\partial_\beta\hat{\pi}_{\mathcal{S}} = -\Delta\hat{H}_{\mathcal{S}}\hat{\pi}^0_{\mathcal{S}} + \gamma^2\partial_\beta(\hat{\pi}_{\mathcal{S}}\hat{X}_{\mathcal{S}}) + \mathcal{O}(\gamma^3). \tag{3.113}$$

Turning to the operator $e^{-\lambda\hat{\pi}_{\mathcal{S}}}$, combining Lemma 3.7.1 and Theorem 3.7.1 leads to

$$e^{-\lambda\hat{\pi}_{\mathcal{S}}} = e^{-\lambda\hat{\pi}^0_{\mathcal{S}}}\big(\hat{\mathbb{I}}_{\mathcal{S}} - \gamma^2\hat{\pi}^0_{\mathcal{S}}\int_0^\lambda d\lambda'\ e^{\lambda'\hat{\pi}^0_{\mathcal{S}}}\hat{X}_{\mathcal{S}}e^{-\lambda'\hat{\pi}^0_{\mathcal{S}}}\big) + \mathcal{O}(\gamma^3). \tag{3.114}$$

Let $\hat{H}_{\mathcal{S}} = \sum_n \epsilon_n |\epsilon_n\rangle\langle\epsilon_n|$ be the eigen-decomposition of the bare Hamiltonian. We first need to evaluate the off-diagonal elements of the operator $\hat{X}_{\mathcal{S}}$ appearing in Theorem 3.7.1 in the energy basis. We first find:

$$(\hat{X}_{\mathcal{S}})_{nm} = \sum_k S_{nk}S_{km}\int_0^\beta d\beta_1\ e^{\beta_1\Delta\epsilon_{kn}}\int_0^{\beta_1} d\beta_2\ e^{\beta_2\Delta\epsilon_{mk}}\Phi_{\mathcal{R}}(-i(\beta_1-\beta_2)) \tag{3.115}$$

where $S_{nk} = \langle\epsilon_n|\hat{V}_{\mathcal{S}}|\epsilon_k\rangle$ and $\Delta\epsilon_{nk} = \epsilon_n - \epsilon_k$. To simplify the above integral we introduce new variables $u = \beta_1 - \beta_2$ and $v = \beta_1 + \beta_2$ and perform the integral over v analytically:

$$(\hat{X}_{\mathcal{S}})_{nm} = \frac{1}{\Delta\epsilon_{nm}}\sum_k D_{nk}S_{km} - D_{mk}S_{kn}, \qquad n \neq m, \tag{3.116}$$

with

$$D_{nk} = S_{nk}\int_0^\beta du\ \Phi_{\mathcal{R}}(-iu)e^{-u\Delta\epsilon_{nk}}, \tag{3.117}$$

For the $n = m$ terms we find

$$(\hat{X}_{\mathcal{S}})_{nn} = \int_0^\beta du\ (\beta - u)\Phi_{\mathcal{R}}(-iu)\bigg(\sum_k (S)^2_{nk} e^{-u\Delta\epsilon_{nk}} - \Phi_{\mathcal{S}}(-iu)\bigg). \tag{3.118}$$

Using this we can compute the SLD with respect to parameter β. Combining (3.113) and (3.114) gives

$$\begin{aligned} e^{-\lambda\hat{\pi}_{\mathcal{S}}}(\partial_\beta\hat{\pi}_{\mathcal{S}})e^{-\lambda\hat{\pi}_{\mathcal{S}}} = &-\Delta\hat{H}_{\mathcal{S}}\hat{\pi}^0_{\mathcal{S}}e^{-2\lambda\hat{\pi}^0_{\mathcal{S}}} + \gamma^2 e^{-\lambda\hat{\pi}^0_{\mathcal{S}}}\partial_\beta(\hat{\pi}^0_{\mathcal{S}}\hat{X}_{\mathcal{S}})e^{-\lambda\hat{\pi}^0_{\mathcal{S}}} \\ &+ \gamma^2\Delta\hat{H}_{\mathcal{S}}(\hat{\pi}^0_{\mathcal{S}})^2\int_0^\lambda d\lambda'\ e^{-(\lambda-\lambda')\hat{\pi}^0_{\mathcal{S}}}\hat{X}_{\mathcal{S}}e^{-(\lambda+\lambda')\hat{\pi}^0_{\mathcal{S}}} \\ &+ \gamma^2\Delta\hat{H}_{\mathcal{S}}(\hat{\pi}^0_{\mathcal{S}})^2 e^{-2\lambda\hat{\pi}^0_{\mathcal{S}}}\int_0^\lambda d\lambda'\ e^{\lambda'\hat{\pi}^0_{\mathcal{S}}}\hat{X}_{\mathcal{S}}e^{-\lambda'\hat{\pi}^0_{\mathcal{S}}} + \mathcal{O}(\gamma^3) \end{aligned} \tag{3.119}$$

We now combine (3.73) with (3.119) and expand the SLD in the energy basis of $\hat{H}_{\mathcal{S}}$:

$$\begin{aligned} \hat{L}_{\mathcal{S}}(\beta) &= 2\int_0^\infty d\lambda\ e^{-\lambda\hat{\pi}_{\mathcal{S}}}(\partial_\beta\hat{\pi}_{\mathcal{S}})e^{-\lambda\hat{\pi}_{\mathcal{S}}}, \\ &= -\Delta\hat{H}_{\mathcal{S}} + \gamma^2\sum_{n,m}\delta\epsilon_n\ \alpha_{n,m}(\beta)\,|\epsilon_n\rangle\langle\epsilon_m| + \mathcal{O}(\gamma^3), \end{aligned} \tag{3.120}$$

where after lengthy but straightforward integration yields the coefficients

$$\alpha_{n,m}(\beta) = \left(\frac{p_n(3p_m + p_n)}{p_m(p_m + p_n)}\right)(\hat{X}_{\mathcal{S}})_{nm} + \left(\frac{2p_n}{p_n + p_m}\right)\partial_\beta(\hat{X}_{\mathcal{S}})_{nm}, \tag{3.121}$$

with $\{p_n\}$ the occupation probabilities of the uncoupled probe $\hat{\pi}^0_{\mathcal{S}}$. Thus so long as we know the bath correlation function $\Phi_{\mathcal{R}}(x)$ and the eigenstates of $\hat{H}_{\mathcal{S}}$, then we can obtain the elements $(\hat{X}_{\mathcal{S}})_{nm}$ from (3.118) and (3.116) and subsequently the SLD (3.120). This formula can be used to determine the optimal measurement needed to obtain the optimal signal-to-noise estimate. The QFI can be found using (3.120), in which case

$$\begin{aligned} \tilde{\mathcal{F}}(\beta) &= \mathrm{tr}_{\mathcal{S}}[\Delta\hat{H}^2_{\mathcal{S}}\hat{\pi}^0_{\mathcal{S}}] - \gamma^2\mathrm{tr}_{\mathcal{S}}[\Delta\hat{H}^2_{\mathcal{S}}\hat{X}_{\mathcal{S}}\hat{\pi}^0_{\mathcal{S}}] - 2\gamma^2\mathrm{tr}_{\mathcal{S}}[\Delta\hat{H}_{\mathcal{S}}\partial_\beta(\hat{\pi}^0_{\mathcal{S}}\hat{X}_{\mathcal{S}})] + \mathcal{O}(\gamma^3), \\ &= \mathrm{tr}_{\mathcal{S}}[\Delta\hat{H}^2_{\mathcal{S}}\hat{\pi}^0_{\mathcal{S}}] + \gamma^2\mathrm{tr}_{\mathcal{S}}[\Delta\hat{H}^2_{\mathcal{S}}\hat{X}_{\mathcal{S}}\hat{\pi}^0_{\mathcal{S}}] - 2\gamma^2\mathrm{tr}_{\mathcal{S}}[\Delta\hat{H}_{\mathcal{S}}(\partial_\beta\hat{X}_{\mathcal{S}})\hat{\pi}^0_{\mathcal{S}}] + \mathcal{O}(\gamma^3), \\ &= \langle\Delta\hat{H}^2_{\mathcal{S}}\rangle + \gamma^2\langle\Delta\hat{H}^2_{\mathcal{S}}\hat{X}_{\mathcal{S}}\rangle - 2\gamma^2\langle\Delta\hat{H}_{\mathcal{S}}(\partial_\beta\hat{X}_{\mathcal{S}}))\rangle + \mathcal{O}(\gamma^3), \end{aligned} \tag{3.122}$$

Note that by comparing (3.118) with the final term in (3.122), we see that only terms that depend on index n contribute to the QFI. This is because $\langle\Delta\hat{H}_{\mathcal{S}}\rangle = 0$. Thus we need to evaluate

$$\frac{\partial}{\partial\beta}\int_0^\beta du\ (\beta - u)\Phi_{\mathcal{R}}(-iu)(S)^2_{nk}e^{-u\Delta\epsilon_{nk}}. \tag{3.123}$$

This turns out to be

$$\frac{\partial}{\partial\beta}\int_0^\beta du\ (\beta-u)\Phi_{\mathcal{R}}(-iu)(S)^2_{nk}e^{-u\Delta\epsilon_{nk}} = \int_0^\beta du\ [(\beta-u)\frac{\partial}{\partial\beta}\Phi_{\mathcal{R}}(-iu) + \Phi_{\mathcal{R}}(-iu)](S)^2_{nk}e^{-u\Delta\epsilon_{nk}}, \tag{3.124}$$

where we made use of the formula from calculus:
$\partial_\beta\left(\int_0^\beta du f(u,\beta)\right) = f(\beta,\beta)+\int_0^\beta du\partial_\beta\left(f(u,\beta)\right)$. Let us now define the following modified auto-correlation function:

$$\Phi'_{\mathcal{R}}(t) := (\beta - it)\Phi_{\mathcal{R}}(t). \tag{3.125}$$

Furthermore, we introduce the operator

$$\hat{S}'_{\mathcal{S}}(t) := e^{it\hat{H}_{\mathcal{S}}}\hat{V}_{\mathcal{S}}e^{-it\hat{H}_{\mathcal{S}}}\hat{V}_{\mathcal{S}}, \tag{3.126}$$

and denote $\Delta\hat{S}'_{\mathcal{S}}(t) = \hat{S}'_{\mathcal{S}}(t) - \langle\hat{S}'_{\mathcal{S}}(t)\rangle$. Substituting (3.118) and (3.124) into (3.122) yields

$$\tilde{\mathcal{F}}(\beta) = \Delta U^2_{\mathcal{S}} + \gamma^2\int_0^\beta du\ \Big[\langle\Delta\hat{H}^2_{\mathcal{S}}\Delta\hat{S}'_{\mathcal{S}}(-iu)\rangle\Phi'_{\mathcal{R}}(-iu) - 2\langle\Delta\hat{H}_{\mathcal{S}}\hat{S}'_{\mathcal{S}}(-iu)\rangle\frac{\partial}{\partial\beta}\Phi'_{\mathcal{R}}(-iu)\Big] + \mathcal{O}(\gamma^3). \tag{3.127}$$

It is straightforward to see that

$$\langle\Delta\hat{H}^2_{\mathcal{S}}\Delta\hat{S}'_{\mathcal{S}}(-iu)\rangle = \frac{\partial^2}{\partial\beta^2}\Phi_{\mathcal{S}}(-iu), \tag{3.128}$$

and

$$\langle\Delta\hat{H}_{\mathcal{S}}\hat{S}'_{\mathcal{S}}(-iu)\rangle = -\frac{\partial}{\partial\beta}\Phi_{\mathcal{S}}(-iu). \tag{3.129}$$

Using this we eventually arrive at an expression for the QFI in terms of the system and bath auto-correlation functions:

$$\tilde{\mathcal{F}}(\beta) = \Delta U^2_{\mathcal{S}} + \gamma^2\int_0^\beta du\ \Big[(\beta-u)\Phi_{\mathcal{R}}(-iu)\frac{\partial^2}{\partial\beta^2}\Phi_{\mathcal{S}}(-iu) + 2\frac{\partial}{\partial\beta}\Phi_{\mathcal{S}}(-iu)\frac{\partial}{\partial\beta}(\beta-u)\Phi_{\mathcal{R}}(-iu)\Big]. \tag{3.130}$$

To conclude the proof we simply use $\tilde{\mathcal{F}}(\beta) = T^4\tilde{\mathcal{F}}(T)$ and $\Delta U^2_{\mathcal{S}} = T^2C_{\mathcal{S}}(T)$, and apply the quantum Crámer-Rao bound (3.65) to parameter T. ■

Theorem 3.7.2 now provides a means of analytically computing the quantum Fisher information for arbitrary strongly-coupled systems up to second order in coupling. In addition, during the proof I also obtained the symmetric logarithmic derivative (SLD) analytically. Recalling Section 3.5, once the SLD is known one can construct the optimal projective measurement that saturates the quantum Crámer-Rao bound (3.65). Therefore Theorem 3.7.2 can be used to find the ultimate accuracy of a coupled quantum thermometer along with the measurement needed to obtain the optimal estimate of temperature.

In order to investigate whether or not strong interactions are beneficial to the thermometer's performance, let us consider measurements of T applied to a bosonic reservoir using a qubit as our probe. This setup can be described using the spin-boson model. Here we set the probe Hamiltonian to be

$$\hat{H}_{\mathcal{S}} = \frac{\epsilon}{2}\hat{\sigma}_z, \tag{3.131}$$

with $\hat{\sigma}_z$ the Pauli spin operator and ϵ the energy gap with $\hbar = 1$. The reservoir Hamiltonian consists of a large collection of harmonic oscillators:

$$\hat{H}_{\mathcal{R}} = \sum_k \omega_k \hat{b}_k^\dagger \hat{b}_k, \tag{3.132}$$

with ω_k the frequency of the k'th oscillator and $(\hat{b}_k^\dagger, \hat{b}_k)$ the respective creation and annihilation operators. Finally, the interaction is chosen so that the probe's x spin component couples to the position of each oscillator:

$$\hat{V}_{\mathcal{S}\cup\mathcal{R}} = \hat{\sigma}_x \otimes \sum_k \left(g_k \hat{b}_k^\dagger + g_k^* \hat{b}_k\right), \tag{3.133}$$

where g_k denotes the coupling strength relative to the overall constant γ. The first function to calculate is the auto-correlation function of $\mathcal{R}$. In the continuum limit this is determined according to

$$\Phi_{\mathcal{R}}(t) = \int_0^\infty d\omega \; J(\omega)\left[\coth\left(\frac{\omega}{2T}\right)\cos(\omega t) - i\sin(\omega t)\right], \tag{3.134}$$

where $J(\omega)$ is the spectral density of $\mathcal{R}$. We choose a generic spectral density that decays exponentially at some large cutoff frequency Ω

$$J_s(\omega) := 2\Omega^{1-s}\omega^s e^{-\omega/\Omega}, \tag{3.135}$$

where $s = 1$ represents an Ohmic bath while $s < 1$ and $s > 1$ are referred to as sub-Ohmic and super-Ohmic respectively. Plugging (3.135) into (3.134) and evaluating the integral leads to the following exact expression:

(3.136)
$$\Phi_{\mathcal{R}}(t) = 2\Omega^{1-s}\beta^{-(s+1)}\Gamma(s+1)\left[\zeta\left(s+1, \frac{1+\beta\Omega - i\Omega t}{\beta\Omega}\right) + \zeta\left(s+1, \frac{1+i\Omega t}{\beta\Omega}\right)\right],$$

where Γ is the Gamma function and $\zeta(z,u)$ is the generalised zeta function. By analytic continuation, we then have

(3.137)
$$\Phi_{\mathcal{R}}(-iu) = 2\Omega^{1-s}\beta^{-(s+1)}\Gamma(s+1)\left[\zeta\left(s+1, \frac{1+\Omega(\beta-u)}{\beta\Omega}\right) + \zeta\left(s+1, \frac{1+\Omega u}{\beta\Omega}\right)\right],$$

To compute the probe correlation function we use the fact that $e^{i\lambda\hat{\sigma}_z} = \cos(\lambda)\hat{\mathbb{I}}_{\mathcal{S}} + i\sin(\lambda)\hat{\sigma}_z$, and hence $e^{it\epsilon\hat{\sigma}_z/2}\hat{\sigma}_x e^{-it\epsilon\hat{\sigma}_z/2}\hat{\sigma}_x = \cos(\epsilon t)\hat{\mathbb{I}}_{\mathcal{S}} + i\sin(\epsilon t)\hat{\sigma}_z$. Similarly one has $\hat{\pi}^0_{\mathcal{S}} = \frac{1}{2}[\hat{\mathbb{I}}_{\mathcal{S}} - \tanh(\beta\epsilon/2)]$. Hence $\Phi_{\mathcal{S}}(t) = \cos(\epsilon t) - i\sin(\epsilon t)\tanh(\beta\epsilon/2)$, and by analytic continuation we get

(3.138)
$$\Phi_{\mathcal{S}}(-iu) = \cosh(\epsilon u) - \sinh(\epsilon u)\tanh(\beta\epsilon/2).$$

It then follows that

(3.139)
$$\frac{\partial}{\partial\beta}\Phi_{\mathcal{S}}(-iu) = \frac{\epsilon}{2}\sinh(\epsilon u)[\tanh^2(\beta\epsilon/2) - 1],$$

(3.140)
$$\frac{\partial^2}{\partial\beta^2}\Phi_{\mathcal{S}}(-iu) = \frac{\epsilon^2}{2}\sinh(\epsilon u)[1 - \tanh^2(\beta\epsilon/2)]\tanh(\beta\epsilon/2).$$

Using Theorem 3.7.2, these analytic expressions can then be used to compute the second-order correction term to the optimal signal-to-noise ratio. In Figure 3.5 we plot the dimensionless correction term $\xi(T)$ for this model for the Ohmic, sub-Ohmic and super-Ohmic reservoir, all as a function of inverse temperature. Clearly the negativity of $\xi(T)$ indicates that the effect of second-order coupling is detrimental to estimating the temperature of $\mathcal{R}$. This suggests that thermometry bounds derived from the weak-coupling assumption $\gamma^2 \ll 1$ may not be achievable in regimes where this assumption no longer holds true. Theorem 3.7.2 provides a tighter constraint on the achievable signal-to-noise ratio for instances where the interaction strength . In fact, for the bosonic reservoir the correction term is valid up to order $\mathcal{O}(\gamma^4)$, given that the interaction term $\sum_k \left(g_k\hat{b}_k^\dagger + g_k^*\hat{b}_k\right)$ is linear in position of each oscillator, which implies that only even values of γ^k in the Taylor expansion survive. This should be compared with the findings in [81],

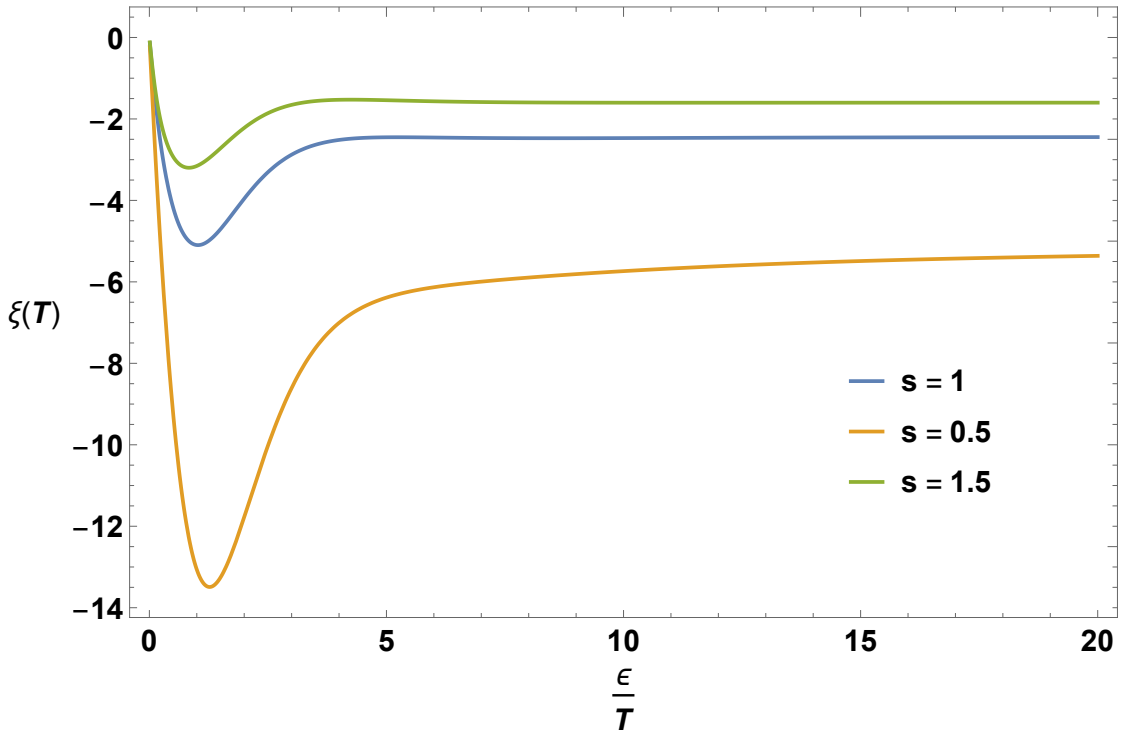


Figure 3.5: **Spin-boson thermometry: High temperature regime.** Plot of the second-order correction term $\xi(T)$ appearing in Theorem 3.7.2 for the spin-boson model as a function of inverse temperature T. Different choices of the Ohmicity parameter s are shown, indicating the effect of coupling on different structured reservoirs. For this high temperature regime we see that the coupling is detrimental to the thermometer's performance.

where it was found that strong coupling can provide enhancements to low temperature measurements in the exactly solvable Caldeira-Leggett model. Since we are neglecting any contributions to the signal-to-noise ratio of order $\mathcal{O}(\gamma^4)$, any improvements must stem from higher order coupling contributions of fourth order or higher. Investigating the overall effect of higher corrections will require going beyond canonical perturbation theory, and is left for future work.

3.8 Concluding remarks and outlook

In this chapter I have shown how non-negligible interactions influence fluctuations in temperature at the nanoscale. To analyse this I introduced a framework for describing strongly-coupled quantum systems centered around the Hamiltonian of mean force. Similar to the formalism in Chapter 2, this approach allows one to assign thermodynamic

potentials to the coupled system that incorporate the effect of the interaction. I then showed that the operator describing the internal energy of the system generally does not commute with the corresponding equilibrium state, which leads to energy coherences and additional temperature-dependent terms. I then explored the effects that this had on the heat capacity of the system. This was shown in Theorem 3.4.1, which is a generalisation of the well-known fluctuation-dissipation relation to systems beyond the weak coupling regime. Here I established a connection between the Wigner-Yanase-Dyson skew information, which is a measure of quantum uncertainty, and the system's heat capacity $\tilde{C}_{\mathcal{S}}$. Proving that the heat capacity, with its strong coupling corrections, vanishes in the zero-temperature limit in accordance with the third law of thermodynamics remains an open question. The appearance of the skew information in Theorem 3.4.1 suggests that quantum coherences may play a role in ensuring its validity. Recent resource-theoretic derivations of the third law [125, 126] could provide a possible avenue for exploring the impact of coherences.

I next explored the effect that strong coupling has on estimations of temperature. Using quantum estimation theory, I considered the quantum Fisher information associated with local temperature estimates made on the system. While the weak-coupling assumption predicts a Heisenberg-like trade-off between temperature and energy uncertainties in (3.82), I found that this relation is no longer valid at arbitrary coupling strengths. Theorem 3.6.1 is a thermodynamic uncertainty relation between energy and temperature extended to all interaction strengths. All fluctuations in energy were found to be related to the effective internal energy operator rather than the bare Hamiltonian of $\mathcal{S}$. This derivation is based on a new bound on the quantum Fisher information for exponential states which was proven in Theorem 3.6.1. As this is valid for any state of full-rank, the bound will be of interest to other areas of quantum metrology. This uncertainty relation shows that for a given finite spread in energy, unbiased estimates of the underlying temperature are limited to a greater extent due to coherences between energy states. These coherences only arise for quantum systems beyond the weak coupling assumption. It was found that these additional temperature fluctuations are quantified by the average Wigner-Yanase-Dyson skew information, thereby establishing a new link between quantum and classical forms of statistical uncertainty in nanoscale thermodynamics. With coherence now understood to be an important resource in the performance of small-scale heat engines [127, 128], these findings suggest that the skew information could be used to unveil further non-classical aspects of quantum thermodynamics. This complements previous results that connect skew information to both unitary phase estimation [129]

and quantum speed limits [88].

By applying the fluctuation-dissipation relation to temperature estimation I derived Theorem 3.6.2, an upper bound on the optimal signal-to-noise ratio expressed in terms of the system's heat capacity. Notably the bound implies that when designing a probe to measure T, its bare Hamiltonian and interaction with the sample should be chosen so as to both maximise $\tilde{C}_{\mathcal{S}}$ whilst minimising the additional dissipation term $\langle \partial_T \hat{E}^*_{\mathcal{S}} \rangle$. It is an interesting open question to consider the form of Hamiltonians that achieve this optimisation in the strong coupling scenario. Furthermore, one expects that improvements to low-temperature thermometry resulting from strong interactions, such as those observed in [81], will be connected to the properties of the effective internal energy operator. In particular, it is clear from Theorem 3.6.2 that any improved scaling of the QFI at low temperatures must be determined by the relative scaling of $\tilde{C}_{\mathcal{S}}$, and exploring this further remains a promising direction of research. Advancements in nanotechnology now enable temperature sensing over microscopic spatial resolutions [130, 131], and understanding how to exploit interactions between a probe and its surroundings will be crucial to the development of these nanoscale thermometers.

The presented approach opens up opportunities for exploring the intermediate regime between the limiting cases [132, 133] of standard thermodynamics with negligible interactions and those where correlations play a prominent role [45, 134, 135]. The results establish a new connection between abstract measures of quantum information theory, such as the quantum Fisher information and skew information, and a material's effective thermodynamic properties. This provides a starting point for future investigations into nanoscale thermodynamics, extending into the regime where the weak coupling assumption is not justified.

In the final section of this chapter, I investigated the impact of coupling strength on the magnitude of the signal-to-noise ratio of a local temperature estimate. While the thermodynamic uncertainty relation offers a novel connection between thermodynamics and quantum skew information, the bound is generally very difficult to calculate. To circumvent this I derived an analytic formula for the signal-to-noise ratio up to second-order in coupling strength, as shown in Theorem 3.7.2. Remarkably, in this regime all quantities can be computed using only knowledge of the sample's spectral density. Furthermore, this result provides analytic expressions for the symmetric logarithmic derivatives, and hence can be used to compute the optimal projective measurement that saturates the bound. This result was then applied to the spin-boson model, with the spin playing the role of the quantum thermometer. It was shown in Figure 3.5 that the

presence of interaction up to third order in coupling is detrimental to the thermometer's accuracy. We conjecture that this behaviour is universal for bosonic reservoirs, and a rigorous proof of this is left for future work. The findings provide tighter constraints on quantum thermometry for when the weak-coupling assumption breaks down. From other investigations on the effect of strong coupling, such as in [81, 136], it has been shown that coupling can provide an enhancement at low temperatures in the Caldeira-Leggett model. However, in [136] this improvement was found to occur at significantly higher coupling strengths of fourth order. The most pressing question left unanswered is whether or not this behaviour is universal, and solving this will require new techniques for representing strongly-coupled systems around absolute zero. Secondly, understanding the connection between the improved scaling of the quantum Fisher information and entanglement is another fundamental avenue to explore.

CHAPTER 4

Quantum work fluctuations in slow processes

An important result in classical stochastic thermodynamics is the fluctuation-dissipation relation (FDR), which states that the dissipated work done along a quasi-static process is proportional to the resulting work fluctuations. To see this, consider a classical system driven in time according to a slow variation in its Hamiltonian, such that it stays close to thermal equilibrium with respect to an external bath at inverse temperature β throughout the process. According to the second law of thermodynamics, the dissipated work along any such process, given by $W_{\text{diss}} = \langle w \rangle - \Delta F$, is always positive. Here $\langle w \rangle$ is the average work done on the system while ΔF is the corresponding change in equilibrium free energy. As we saw in Chapter 2, outside of the thermodynamic limit the work will fluctuate with variance $\sigma_w^2 = \langle w^2 \rangle - \langle w \rangle^2$, and linear response theory tells us that these fluctuations are responsible for the excess dissipation [18, 54, 137, 138]:

$$W_{\text{diss}} = \frac{1}{2}\beta\sigma_w^2 \geq 0. \tag{4.1}$$

This work fluctuation-dissipation relation (FDR) is a consequence of the fact that the corresponding work distribution $p(w)$ becomes Gaussian in the quasi-static limit, which has been rigorously proven in [55]. One important consequence of (4.1) is that any optimal slow process that minimises dissipation will subsequently minimise the fluctuations. Such an optimisation can be obtained using techniques from differential geometry [139]. In this situation one can equip the thermodynamic state space with a unique Riemannian metric [140, 141], and optimal protocols can be found by calculating the associated geodesics through this space [19, 142]. Interestingly, the metric over the set of thermal states is

identical to the classical Fisher information (3.60) with respect to time [19].

In the quantum regime, work is typically defined via the two-projective-measurement (TPM) scheme [16, 143]. For a system in contact with a thermal bath, the work statistics are obtained from measurements of the total energy changes as the system is driven away from equilibrium. In this framework, the standard fluctuation relations are satisfied and one recovers the usual laws of thermodynamics at the stochastic level despite the presence of quantum coherence generated along a typical non-equilibrium process [6, 17]. Naively, much like the classical case one might expect a Gaussian work distribution for sufficiently slow driving, where the system stays close to an instantaneous thermal state throughout the protocol. In this chapter I will show that this is no longer true in the quantum regime. Moreover, a general protocol necessarily leads to a non-Gaussian work distribution due to the non-commutativity between the system's Hamiltonian and its associated power operator. To show this I will consider the work fluctuations in a slowly-driven open quantum system and derive a modified work FDR. It will be seen that this result differs from (4.1) through an additional contribution that is related to the Wigner-Yanase-Dyson skew information (3.15), which quantifies the amount of non-classical power fluctuations generated along the process. More specifically, this modified work FDR will imply that quantum work fluctuations actually exceed the amount of dissipation. Crucially, the quantum corrections to the FDR and non-Gaussian shape of the work distribution can be witnessed experimentally via the TPM work statistics [144–146]. Thus these results provide a direct means of detecting non-classicality in quantum non-equilibrium processes.

Finally, I prove that these fluctuations are related to a Riemannian metric over the space of thermal states. While it has been shown that the dissipated work is also related to a metric over the thermal states [147], we will see that there exists an in-equivalence between these two geometric quantities that only vanishes in the classical commutative regime. This surprising result suggests that the standard geometric structure of thermodynamics changes when moving to the quantum regime. These in-equivalent metrics over the manifold fundamentally rule out optimal protocols that simultaneously minimise both W_{diss} and σ_w. However, the framework derived in this chapter can be used to find optimal trade-offs between dissipation and fluctuations through computations of the different geodesics through the thermodynamic state space. This chapter is based on an unpublished manuscript "Work fluctuations in slow processes: quantum signatures and optimal control" that has been done in collaboration with Matteo Scandi, Janet Anders and Marti Perarnau-Llobet.

4.1 Thermodynamic geometry in classical systems

As was shown in Chapter 2, a classical open system driven away from equilibrium satisfies the Jarzynski fluctuation relation, which relates the stochastic work done along the non-equilibrium process to the change in equilibrium free energy. As first observed by Jarzynski in [18], this equality implies a form of work fluctuation-dissipation relation if the work distribution is Gaussian. More specifically, if $p(w) \propto \exp((w - \langle w \rangle)^2 / 2\sigma_w^2)$, then we get the following implication:

$$\left\langle e^{-\beta w} \right\rangle = e^{-\beta \Delta F} \Longrightarrow W_{\text{diss}} = \frac{1}{2} \beta \sigma_w^2. \tag{4.2}$$

This follows from the fact that a Gaussian distribution has vanishing cumulants beyond the average and variance. The relation takes the same form as the many fluctuation-dissipation relations encountered in physics. Namely, it implies that any additional work dissipated in to the environment due to deviations from equilibrium results in a proportional stochastic fluctuation of the same quantity, and vice-versa.

As was first suggested in [18], one should expect a Gaussian work distribution to emerge when the system Hamiltonian is driven very slowly. This assumption can be loosely justified by appealing to the central limit theorem [54]. In order to rigorously prove this assumption, Seifert and Speck explored the asymptotic behaviour of the work distribution in the long time limit [55]. They begin by considering a general open classical system connected to a thermal environment at temperature $k_B T = 1/\beta$, where the system Hamiltonian $H_\lambda(x)$ is driven in time according to some protocol λ_t. Here we label the system coordinates by the vector $x = (x_0, x_1, ...)$. The system is assumed to obey a generic time-dependent Fokker-Planck equation of the form

$$\partial_t \rho_t = \mathcal{L}_\lambda[\rho_t] = \left[\frac{\partial}{\partial x_i} \mu_{ij} \left(\frac{\partial H_\lambda}{\partial x_j} + \frac{1}{\beta} \frac{\partial}{\partial x_j} \right) \right] \rho_t, \tag{4.3}$$

where μ_{ij} are the mobility coefficients. In addition it is also assumed that the stationary solution to (4.3) for any fixed value of control parameter λ is a corresponding Gibbs state:

$$\mathcal{L}_\lambda[\pi_\lambda] = 0 \Longrightarrow \pi_\lambda := \frac{e^{-\beta H_\lambda}}{Z_\lambda}. \tag{4.4}$$

In other words, as the system evolves in time the reservoir tends to bring it back to an instantaneous equilibrium state. Clearly if the process is slow enough the system will remain in equilibrium and form a perfectly quasi-static process with no dissipation or fluctuations. However, for finite time there will be small deviations from equilibrium.

The fluctuating work up to time τ is then given by a integral over a particular phase space trajectory:

$$w[x(\tau),\tau] := \int_0^\tau dt' \ \dot{\lambda}\frac{\partial H_\lambda}{\partial \lambda}(x(t')), \tag{4.5}$$

If one considers the stochastic pair of variables $\{x, w\}$ and their joint probability $p(x, w)$, this distribution also satisfies a Fokker-Planck equation of the form

$$\partial_t p_t := \left[\mathcal{L}_\lambda - \dot{\lambda}\frac{\partial H_\lambda}{\partial \lambda}\frac{\partial}{\partial w}\right]p_t, \tag{4.6}$$

with initial condition $p_0 = \pi_{\lambda(0)}\delta(w)$. At any given time the work distribution can then be obtained by simply integrating the solution to (4.6) over x:

$$p(w) := \int dx \ p(x, w). \tag{4.7}$$

By introducing a projection-operator technique, the authors in [55] then expand the general solution to (4.6) in powers of the total duration of the protocol τ. Remarkably, one indeed finds that the solution to (4.6), after integrating over x leads to a Gaussian work distribution up to first order in $1/\tau$, ie.

$$p(w) = \frac{1}{\sqrt{2\pi\sigma_w^2}}e^{(w-\langle w\rangle)^2/2\sigma_w} + \mathcal{O}(1/\tau^2). \tag{4.8}$$

Returning to (4.2), this rigorously proves the work fluctuation-dissipation relation for slow processes. Note that while we have assumed that the stationary solution to the open system dynamics is Gibbsian, the result may be extended to describe arbitrary non-equilibrium steady-states [148]. In this case the fluctuation-dissipation relation relates the stochastic entropy production to its variance.

The work FDR (4.1) is a fundamental result as it demonstrates that an efficient protocol with minimal dissipation will also be the most reliable, namely with minimal work fluctuations. This is crucial for designing optimal molecular machines that operate in finite time. Thus in classical stochastic thermodynamics optimal slow protocols can be found by simply looking for the most efficient process at a given duration τ. The solution to such an optimisation problem will then be guaranteed to have minimal unwanted fluctuations. It should be noted that when constructing faster processes, say beyond first order in $1/\tau$, one will naturally have to compromise between minimising either dissipation or fluctuations. A typical example of this trade-off can be found in [149]. In addition, faster processes will deviate further from the efficiency allowed by the second law, and so

throughout this chapter I will only focus on slow processes up to order $1/\tau$, since these processes are fundamentally the most efficient allowed in finite time.

The problem of finding an optimally efficient process can be solved through the use of differential geometry [139]. It was first recognised in the 1980's that macroscopic thermodynamics shares a close connection with Riemannian geometry. In particular, one may relate the thermodynamic free energy functional to a metric over the manifold of equilibrium states, parameterised by pressure, volume and other external controls attributed to the system [140, 141]. Through knowledge of the metric one can then apply techniques from differential geometry to compute geodesics through this equilibrium manifold. These geodesics, which represent the shortest curves joining an initial and final equilibrium configuration, provide an explicit description of the optimal process with minimal entropy production or dissipated work. While the original results on thermodynamic geometry were based on macroscopic thermodynamics and the assumption of endoreversibility, this framework has also been generalised to describe small scale systems governed by stochastic thermodynamics [19, 142]. Let $\vec{\lambda}_t = (\lambda_1(t), \lambda_2(t), ...)$ represent the external control parameters of the system Hamiltonian. Remarkably, Crooks has shown that the average dissipated work done along a slow process admits the following geometric expression:

$$W_{\text{diss}} = \int_0^\tau dt \, \left[\frac{d\vec{\lambda}}{dt}\right]^T . \mathbf{g}(\vec{\lambda}) . \left[\frac{d\vec{\lambda}}{dt}\right], \tag{4.9}$$

where we have neglected terms of order $\mathcal{O}(1/\tau^2)$. Here $\mathbf{g}(\lambda_t)$ is a positive, symmetric metric tensor exactly given by the Fisher-Rao information matrix:

$$g_{ij}(\vec{\lambda}) := \left\langle \frac{\partial \ln \pi_\lambda}{\partial \lambda_i} . \frac{\partial \ln \pi_\lambda}{\partial \lambda_j} \right\rangle, \tag{4.10}$$

where the average is taken with respect to the equilibrium state π_λ. In classical information geometry, the Fisher-Rao metric represents the unique contractive metric over the space of probability distributions [139]. Here we see that when considering the subset of thermal states, namely the equilibrium manifold, the Fisher-Rao metric controls how much dissipation is produced when moving between different states in the manifold. We remark here that the reservoir is assumed to be perfectly thermalising, though the result can be extended to more general open systems [142].

By the geometric structure of (4.9) we see that the dissipated work is essentially an action functional of the metric $\mathbf{g}(\lambda)$. In addition to the action one can construct a length

over the equilibrium manifold, given by

$$\ell(\vec{\lambda}_0, \vec{\lambda}_\tau) := \int_0^\tau dt \, \sqrt{\left[\frac{d\vec{\lambda}}{dt}\right]^T . \mathbf{g}(\vec{\lambda}) . \left[\frac{d\vec{\lambda}}{dt}\right]}, \tag{4.11}$$

This quantity is often referred to as the *thermodynamic length* [19], since the associated geodesic $\ell^*(\vec{\lambda}_0, \vec{\lambda}_\tau) = \inf \ell(\vec{\lambda}_0, \vec{\lambda}_\tau)$ given by the infinimum over all curves with the same boundary points, is a positive, symmetric quantity satisfying the triangle inequality:

$$\ell^*(\vec{\lambda}_a, \vec{\lambda}_b) \leq \ell^*(\vec{\lambda}_a, \vec{\lambda}_c) + \ell^*(\vec{\lambda}_c, \vec{\lambda}_b) \tag{4.12}$$

and hence forms a proper notion of distance over the equilibrium manifold. Notably it does not depend on the duration of the protocol τ. If follows from the Cauchy-Schwarz inequality that the dissipated work and thermodynamic length are related according to

$$W_{\text{diss}} \geq \ell^2/\tau. \tag{4.13}$$

Finally, in order to find the optimal protocol we apply the Euler-Lagrange equations to the action functional (4.9). For a single parameter, say λ_t, this amounts to solving the equation

$$\frac{d\mathcal{C}}{d\lambda} = \frac{d}{dt}\left[\frac{\partial f}{\partial \dot{\lambda}}\right], \tag{4.14}$$

where the cost function is the metric expression $\mathcal{C}(\lambda, \dot{\lambda}) = \dot{\lambda}^2 g(\lambda)$. The solution turns out to be given by [142]

$$\dot{\lambda}_t^{opt} := \frac{(\lambda_\tau - \lambda_0) g^{-1/2}(\lambda_t)}{\int_0^\tau dt \; g^{-1/2}(\lambda_t)}. \tag{4.15}$$

This result implies that the optimal process with minimal dissipation with a velocity that is inversely proportional to the square root of the Fisher-Rao metric. The optimisation problem can be extended straightforwardly to the multi-parameter case. The curve $\dot{\lambda}_t^{opt}$ also defines the geodesic through the equilibrium manifold, since the curve of minimal action is also the shortest. As a result, the optimal solution saturates the geometric lower bound (4.13).

To summarise this section, we have provided an overview of the geometric approach to classical thermodynamics. The above procedure can be used to design optimally efficient stochastic processes with minimal dissipation, as long as the process is slow enough. It follows from the work fluctuation-dissipation relation (4.1) that the optimal curve $\dot{\lambda}_t^{opt}$ will also minimise the fluctuations in work. Throughout the rest of this chapter we will investigate the extension of both the work FDR and the connection between thermodynamics and geometry for quantum systems. It will be shown that a variety of new phenomena emerge, contrasting with the results found in this section.

4.2 The quantum definition of work

As we saw in Chapter 2, the fluctuating work done on an evolving classical system is given by the integrated rate of energy change over a given trajectory in phase space. However, in the quantum regime such trajectories no longer exist due to the non-commutativity between position and momentum. Alternatively, work can be defined through measurements of the system's energy. For closed quantum systems undergoing unitary dynamics the fluctuating work can be identified within the Two-Projective Measurement approach (TPM) [16, 143]. Here the energy of the system is measured at the beginning and end of the system's evolution and the work, w, is taken to be the difference between these two energy measurements. The two-point definition of work has been used to derive quantum fluctuation theorems [17, 150–152], which have been tested experimentally [146, 153]. The TPM scheme has also been generalised to open quantum systems [36, 154]. The choice of the TPM work definition has been supported by showing its direct correspondence with the classical definition of work in the semi-classical limit [155].

In order to understand how to construct the definition of quantum work, we consider a composite system and reservoir $\mathcal{S} \cup \mathcal{R}$ evolving unitarily. Assume that the total Hamiltonian of $\mathcal{S}$ is driven locally with respect to $\mathcal{S}$, with

$$\hat{H}_{\mathcal{S}\cup\mathcal{R}}(t) := \hat{H}_{\mathcal{S}}(t) + \hat{H}_{\mathcal{R}} + \gamma \hat{V}_{\mathcal{S}\cup\mathcal{R}}, \tag{4.16}$$

The process takes place across some time interval $t \in [0, \tau]$, and the initial state of $\mathcal{S} \cup \mathcal{R}$ is assumed to be in a global canonical equilibrium state. Due to the change in the total Hamiltonian, work is done on the system and it is driven away from equilibrium. The resulting unitary for driving of the Hamiltonian is given by

$$\hat{U}(0,\tau) := \overleftarrow{\mathcal{T}} \exp\left[-\frac{i}{\hbar} \int_0^\tau dt\ \hat{H}_{\mathcal{S}\cup\mathcal{R}}(t) \right]. \tag{4.17}$$

Furthermore let us denote the instantaneous spectral decomposition of the total Hamiltonian by

$$\hat{H}_{\mathcal{S}\cup\mathcal{R}}(t) = \sum_n \epsilon_m(t) \left|\epsilon_m(t)\right\rangle \left\langle\epsilon_m(t)\right|. \tag{4.18}$$

For a single run of the process, the work done on the system is a fluctuating variable and is given by the change in total energy after projecting onto the energy basis of the joint system and reservoir at times $t = 0$ and $t = \tau$ [16]. The resulting work probability

distribution is thus given by the following:

$$p(w) := \sum_{n,m} \delta[w - \epsilon_m(\tau) + \epsilon_n(0)] p[\epsilon_m(\tau)|\epsilon_n(0)] p[\epsilon_n(0)], \tag{4.19}$$

where

$$\begin{aligned} p[\epsilon_n(0)] &:= \langle \epsilon_n(0)| \, \hat{\pi}_{\mathcal{S}\cup\mathcal{R}}(0) \, |\epsilon_n(0)\rangle \, , \\ p[\epsilon_m(\tau)|\epsilon_n(0)] &:= \left| \, \langle \epsilon_m(\tau)| \, \hat{U}_{0,\tau} \, |\epsilon_m(0)\rangle \, \right|^2 . \end{aligned} \tag{4.20}$$

are the resulting conditional probabilities obtained via the Born rule. The statistics of work can be conveniently described using the moment-generating function, which is defined as the Fourier transform of the work distribution; $G(u) = \int du \; p(w) e^{iuw}$. From (4.19) one finds

$$G(u) := \mathrm{tr}\left[e^{iu\hat{H}^H_{\mathcal{S}\cup\mathcal{R}}(\tau)} e^{-iu\hat{H}_{\mathcal{S}\cup\mathcal{R}}(0)} \hat{\pi}_{\mathcal{S}\cup\mathcal{R}}(0) \right], \tag{4.21}$$

where we denote the Hamiltonian $\hat{H}^H_{\mathcal{S}\cup\mathcal{R}}(t) = U^\dagger(t,0)\hat{H}_{\mathcal{S}\cup\mathcal{R}}(t)U(t,0)$ in the Heisenberg picture. The work moments can then be obtained via the equation

$$\langle w^k \rangle = (-i)^k \frac{d^k}{du^k} G(u) \bigg|_{u=0} . \tag{4.22}$$

It can be shown that $G(u)$ is an analytic function on the strip $\mathscr{S} := \{u \mid 0 \leq \mathrm{Im}(u) \leq \beta, -\infty < \mathrm{Re}(u) < \infty\}$ [150], and by setting $u = i\beta$ in (4.21) one obtains the quantum version of the Jarzynski equality:

$$\Delta F_{\mathcal{S}} = -\frac{1}{\beta} \ln \left\langle e^{-\beta w} \right\rangle, \tag{4.23}$$

with $\Delta F_{\mathcal{S}}$ the change in equilibrium free energy between $t = 0$ and $t = \tau$. It follows from Jensen's inequality that the dissipated work is positive:

$$W_{\mathrm{diss}} = \langle w \rangle - \Delta F \geq 0, \tag{4.24}$$

which is the quantum version of the second law of thermodynamics.

An important observable that will be the focus of this chapter is the power operator [156], which is defined as the rate of change in energy:

$$\hat{P}_{\mathcal{S}}(t) := \frac{\partial}{\partial t} \hat{H}_{\mathcal{S}}(t). \tag{4.25}$$

Importantly, this operator does not necessarily commute with the system Hamiltonian, unlike in classical thermodynamics. It will be convenient to express the work moments

in terms of this power operator. To do this we now introduce the following nested commutator:

$$\mathcal{C}_0[\hat{A},\hat{B}] := \hat{B}; \quad \mathcal{C}_{j+1}[\hat{A},\hat{B}] := [\hat{A},\mathcal{C}_j[\hat{A},\hat{B}]], \tag{4.26}$$

and denote $\mathcal{C}_j^H[\hat{A},\hat{B}] = \hat{U}^\dagger(t,0)\mathcal{C}_j[\hat{A},\hat{B}]\hat{U}(t,0)$ as the same operator in the Heisenberg picture, along with $\hat{P}_{\mathcal{S}}^H(t) = \hat{U}^\dagger(t,0)\hat{P}_{\mathcal{S}}(t)\hat{U}(t,0)$ as the power operator in the Heisenberg picture. It is shown in [154] that the moment generating function for the the work distribution has the following series expansion:

$$\begin{aligned} G(u) = \mathrm{tr}\Bigg[\overrightarrow{\mathcal{T}}\exp\Bigg(\int_0^\tau dt\ \sum_{j=1}^{\infty}(-1)^{j+1}\frac{(iu)^j}{j!2^j}\mathcal{C}_j^H[\hat{H}_{\mathcal{S}\cup\mathcal{R}}(t),\hat{P}_{\mathcal{S}}(t)]\Bigg) \\ \times\overleftarrow{\mathcal{T}}\exp\Bigg(\int_0^\tau dt\ \sum_{j=1}^{\infty}\frac{(iu)^j}{j!2^j}\mathcal{C}_j^H[\hat{H}_{\mathcal{S}\cup\mathcal{R}}(t),\hat{P}_{\mathcal{S}}(t)]\Bigg)\hat{\pi}_{\mathcal{S}\cup\mathcal{R}}(0)\Bigg], \end{aligned} \tag{4.27}$$

with $\overrightarrow{\mathcal{T}}(\overleftarrow{\mathcal{T}})$ the (anti)-time ordering operator. It is then straightforward to show that the first two moments of (4.27) are given as follows:

$$\langle w\rangle = \int_0^\tau dt\ \mathrm{tr}_{\mathcal{S}}\big[\hat{P}_{\mathcal{S}}(t)\rho_{\mathcal{S}}(t)\big], \tag{4.28}$$

$$\langle w^2\rangle = \int_0^\tau dt\int_0^\tau dt'\ \mathrm{Re}\left(\mathrm{tr}\big[\hat{P}_{\mathcal{S}}^H(t')\hat{P}_{\mathcal{S}}^H(t)\hat{\pi}_{\mathcal{S}\cup\mathcal{R}}(0)\big]\right). \tag{4.29}$$

We note here that while the average work depends only on the reduced system state $\rho_{\mathcal{S}}(t) = \mathrm{tr}_{\mathcal{R}}[\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)]$, the second moment generally depends on the evolution through the full Hilbert space due to correlations between system and reservoir.

These pair of equations provide a useful link between the work done and the integrated power over time. However, in general one will not have access to the full unitary dynamics of $\mathcal{S}\cup\mathcal{R}$. It is most often the case that only the system dynamics are known via some Markovian master equation. Here we will assume that the equation of motion for the reduced state of $\mathcal{S}$ is of Lindblad form [123]:

$$\frac{\partial}{\partial t}\hat{\rho}_{\mathcal{S}}(t) := \mathscr{L}_t[\hat{\rho}_{\mathcal{S}}(t)], \tag{4.30}$$

with

$$\mathscr{L}_t[(.)] := \theta_t(.) + \mathscr{D}_t(.) \tag{4.31}$$

Here we divide the Lindbladian into a unitary and dissipative part respectively.

An important assumption leading to the validity of (4.30) is the so-called Born-Markov approximation [157], which assumes that the interaction term $\hat{V}_{\mathcal{S}\cup\mathcal{R}}$ is sufficiently weak so that the total evolution of $\mathcal{S}\cup\mathcal{R}$ is essentially uncorrelated:

$$\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t) \simeq \hat{\rho}_{\mathcal{S}}(t)\otimes\hat{\pi}_{\mathcal{R}}, \tag{4.32}$$

with $\hat{\pi}_{\mathcal{R}}$ a fixed equilibrium state of $\mathcal{R}$. With this additional assumption, it is possible to express the second work moment (4.29) in terms of the system degrees of freedom. This is summarised in the following lemma:

Lemma 4.2.1. *Suppose the system evolves according to the Lindblad equation* (4.30) *and satisfies the Born-Markov assumption* (4.32)*. Then the second work moment is given by*

$$\langle w^2\rangle = \int_0^\tau dt\int_0^\tau dt'\ \mathrm{Re}\left(\mathrm{tr}_S\big[\hat{P}_{\mathcal{S}}(t')\ \overleftarrow{P}(t',t)[\hat{P}_{\mathcal{S}}(t)\hat{\rho}_{\mathcal{S}}(t)]\big]\right), \tag{4.33}$$

where

$$\overleftarrow{P}(t_1,t_2) = \overleftarrow{\mathcal{T}}\exp\left(\int_{t_2}^{t_1} d\nu\ \mathscr{L}_\nu\right), \tag{4.34}$$

is the propagator for the Liouvillian in (4.30)*.*

Proof. Let us now consider two hermitian operators $\hat{A}_{\mathcal{S}},\hat{B}_{\mathcal{S}}\in\mathcal{S}$ acting on the system. In the Heisenberg picture these operators are expressed as $\hat{X}^H_{\mathcal{S}}(t)=\hat{U}^\dagger(t,0)\hat{X}_{\mathcal{S}}\hat{U}(t,0)$ with $X=A,B$. We also write the evolved density operator for the composite state by $\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)=\hat{U}(t,0)\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(0)\hat{U}^\dagger(t,0)$ with initial condition $\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)=\hat{\pi}_{\mathcal{S}\cup\mathcal{R}}(0)$. We are concerned with evaluating the two-time correlation function $\langle\hat{A}^H_{\mathcal{S}}(t')\hat{B}^H_{\mathcal{S}}(t)\rangle$ with $t'\geq t$, which can be expressed as follows:

$$\begin{aligned}
\langle\hat{A}^H_{\mathcal{S}}(t')\hat{B}^H_{\mathcal{S}}(t)\rangle &= \mathrm{tr}[\hat{A}^H_{\mathcal{S}}(t')\hat{B}^H_{\mathcal{S}}(t)\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(0)],\\
&= \mathrm{tr}[\hat{U}^\dagger(t',0)\hat{A}_{\mathcal{S}}\hat{U}(t',0)\hat{U}^\dagger(t,0)\hat{B}_{\mathcal{S}}\hat{U}(t,0)\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(0)],\\
&= \mathrm{tr}[\hat{A}_{\mathcal{S}}\hat{U}(t',t)\hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)\hat{U}^\dagger(t',t)],\\
&= \mathrm{tr}_{\mathcal{S}}\big[\hat{A}_{\mathcal{S}}\ \mathrm{tr}_{\mathcal{R}}[\hat{U}(t',t)\hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)\hat{U}^\dagger(t',t)]\big],
\end{aligned} \tag{4.35}$$

where in the final line we used the fact that $\hat{B}_{\mathcal{S}}=\hat{B}_{\mathcal{S}}\otimes\hat{\mathbb{I}}_{\mathcal{R}}$. Setting $\tilde{t}=t'-t\geq 0$, a simple change in variables gives

$$\langle\hat{A}^H_{\mathcal{S}}(t+\tilde{t})\hat{B}^H_{\mathcal{S}}(t)\rangle = \mathrm{tr}_{\mathcal{S}}\big[\hat{A}_{\mathcal{S}}\ \hat{\chi}_{\mathcal{S}}(\tilde{t})\big], \tag{4.36}$$

where

$$\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(\tilde{t}) = \hat{U}(t+\tilde{t},t)\hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)\hat{U}^{\dagger}(t+\tilde{t},t),$$
$$\hat{\chi}_{\mathcal{S}}(\tilde{t}) = \mathrm{tr}_{\mathcal{R}}[\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(\tilde{t})]. \tag{4.37}$$

Now observe that $\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(\tilde{t})$ is the solution to the following equation of motion:

$$\frac{d}{d\tilde{t}}\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(\tilde{t}) := -\frac{i}{\hbar}[\hat{H}_{\mathcal{S}\cup\mathcal{R}}(\tilde{t}), \hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(\tilde{t})], \tag{4.38}$$

with initial condition $\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(0) = \hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(0)$. We now use the Born approximation (4.32), which implies that initial condition to (4.38) factorises according to $\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(0) = \hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}}(t)\otimes\hat{\pi}_{\mathcal{R}}$. Given that the initial operator $\hat{\chi}_{\mathcal{S}\cup\mathcal{R}}(0)$ here factorises and obeys the same global equation of motion given by (4.38) with respect to $\tilde{t}$ as the state $\hat{\rho}_{\mathcal{S}\cup\mathcal{R}}(t)$, we obtain the following solution after tracing out the bath degrees of freedom:

$$\hat{\chi}_{\mathcal{S}}(\tilde{t}) = \overleftarrow{P}(t+\tilde{t},t)[\hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}}(t)], \tag{4.39}$$

Combining this with (4.36) we have

$$\langle \hat{A}^H_{\mathcal{S}}(t+\tilde{t})\hat{B}^H_{\mathcal{S}}(t)\rangle = \mathrm{tr}_S\big[\hat{A}_{\mathcal{S}}\overleftarrow{P}(t+\tilde{t},t)[\hat{B}_{\mathcal{S}}\hat{\rho}_{\mathcal{S}}(t)]\big]. \tag{4.40}$$

We now identify $t+\tilde{t} = t'$ inside the integral in (4.29), which is valid considering that the integrand is symmetric with respect to interchanging $t \to t'$. Furthermore we set $\hat{A}_{\mathcal{S}} = \hat{P}_{\mathcal{S}}(t')$ and $\hat{B}_{\mathcal{S}} = \hat{P}_{\mathcal{S}}(t)$. Combining this all together gives us the expression for $\langle w^2\rangle$ in terms of the system degrees of freedom. ■

Lemma 4.2.1 demonstrates that the work fluctuations can be computed from the reduced system dynamics alone, in contrast to the higher order work moments that depend on the reservoir degrees of freedom. This fact will be important when considering the behaviour of these fluctuations in the slow-driving limit.

4.3 The quantum work fluctuation-dissipation relation

Now that we have expressions for the average work and fluctuations, it remains to be answered whether or not the FDR (4.1) holds in the quantum regime. Similarly to the classical construction, we will now assume that the stationary state of the system evolution (4.30) is a local Gibbs state. To simplify notation we will write $\hat{H}_{\mathcal{S}}(t) = \hat{H}_t$,

$(\partial/\partial t)\hat{H}_{\mathcal{S}}(t) = \hat{P}_t$, $\hat{\rho}_{\mathcal{S}}(t) = \hat{\rho}_t$ and $\hat{\pi}_{\mathcal{S}}(t) = \hat{\pi}_t$. To ensure that the thermal state is a stationary solution of (4.30), it is assumed that the Lindbladian $\mathscr{L}_t$ has only negative real eigenvalues along with a unique zero-eigenvalue. In this case the resulting evolution map is said to be mixing [123]. For this mixing map the reservoir tends to keep the system close to an instantaneous thermal state, such that for any normalised state $\hat{\rho}$,

$$\lim_{\nu\to\infty} e^{\nu\mathscr{L}_t}[\hat{\rho}] = \hat{\pi}_t, \quad \forall t. \tag{4.41}$$

To further characterise the open system dynamics and ensure that the Lindbladian has an instantaneous thermal fixed point, we next specify the structure of the dissipator in (4.30). For weakly-coupled open systems, such as the ubiquitous Davies maps used throughout quantum thermodynamics [123], the vast majority of open quantum systems take the following form [158, 159]:

$$\theta_t(.) := -i[H_t, (.)], \tag{4.42}$$

$$\mathscr{D}_t(.) := \sum_{\omega_t}\sum_{\alpha,\alpha'} \gamma_{\alpha\alpha'}(\omega_t)\bigg(\hat{A}_{\alpha'}(\omega_t)(.)\hat{A}^\dagger_{\alpha}(\omega_t) - \frac{1}{2}\{\hat{A}^\dagger_{\alpha}(\omega_t)\hat{A}_{\alpha'}(\omega_t), (.)\}\bigg), \tag{4.43}$$

where $\hat{A}_\alpha(\omega_t)$ denote the set of eigenoperators with ω_t the modes of the system Hamiltonian. They satisfy the following symmetry property:

$$\hat{A}^\dagger_\alpha(\omega_t) = \hat{A}_\alpha(-\omega_t). \tag{4.44}$$

The term $\gamma_{\alpha\alpha'}(\omega_t)$ is a positive semi-definite matrix given by the Fourier transform of the bath correlation function, and satisfies the so-called KMS condition:

$$\gamma_{\alpha\alpha'}(-\omega_t) = \gamma_{\alpha'\alpha}e^{-\beta\omega_t}(\omega_t). \tag{4.45}$$

An important property of these Lindbladians is the condition of *detailed balance*, which constrains the form of the eigenoperators [158]:

$$\begin{aligned} \hat{\pi}_t\ \hat{A}_\alpha(\omega_t) &= e^{\beta\omega_t}\hat{A}_\alpha(\omega_t)\hat{\pi}_t, \\ \hat{\pi}_t\ \hat{A}^\dagger_\alpha(\omega_t) &= e^{-\beta\omega_t}\hat{A}^\dagger_\alpha(\omega_t)\hat{\pi}_t. \end{aligned} \tag{4.46}$$

In addition, the unitary and dissipative parts commute:

$$[\theta_t(.), \mathscr{D}_t(.)] = [\theta_t(.), \mathscr{D}^\dagger_t(.)] = 0. \tag{4.47}$$

Physically, detailed balance implies that the probability of a forward transition from equilibrium is equal to the time-reversed transition.

It is assumed that the changes in $\hat{H}_t$ are much slower than the thermalisation timescale of the dissipative evolution. If one expands the state of $\mathcal{S}$ in terms of the inverse time-scale of the protocol, then the system is approximately given by its instantaneous thermal state plus a correction term of order $1/\tau$. To see this, we will need to define the Drazin inverse of the Lindbladian, $\mathscr{L}_t^+$ [160]. This operator can be thought of as a generalised inverse of $\mathscr{L}_t$ which satisfies the following conditions:

$$\begin{aligned}
&(i) \quad \mathscr{L}_t\mathscr{L}_t^+[\hat{A}] = \mathscr{L}_t^+\mathscr{L}_t[\hat{A}] = \hat{A} - \hat{\pi}_t \mathrm{tr}[\hat{A}], \\
&(ii) \quad \mathscr{L}_t^+[\hat{\pi}_t] = 0, \\
&(iii) \quad \mathrm{tr}[\mathscr{L}_t^+[\hat{A}]] = 0.
\end{aligned}$$

Now we apply this superoperator to both sides of the Lindblad equation (4.30), and a straightforward rearrangement leads to the following expression [161]:

$$\hat{\rho}_t = \left[\mathbb{I} - \mathscr{L}_t^+ \frac{d}{dt}\right]^{-1} \hat{\pi}_t, \tag{4.48}$$

Using the geometric series $(1-x)^{-1} = 1 + x + x^2 + ...$ and Lemma 3.4.1, we get the following perturbation expansion for the system state, to first order in $1/\tau$:

$$\hat{\rho}_t = \hat{\pi}_t + \mathscr{L}_t^+ \mathbb{J}_{\hat{\pi}_t}(\dot{\hat{P}}_t) + \mathcal{O}(1/\tau^2) \tag{4.49}$$

where we have defined the linear operator

$$\mathbb{J}_{\hat{\varrho}}(\hat{A}) := \int_0^1 da\ \hat{\varrho}^a \Delta_{\hat{\varrho}} \hat{A} \hat{\varrho}^{1-a}, \tag{4.50}$$

with $\Delta_{\hat{\varrho}}\hat{A} = \hat{A} - \mathrm{tr}[\hat{A}\hat{\varrho}]$. Note that throughout this work we consider fixed boundary conditions on the system Hamiltonian. In this sense the factor $1/\tau$ represents the relative speed of the protocol, and small values indicate slow driving. Now that we have a first order correction term analytically, this can be used to derive an expression for the dissipated work (4.28) up to first order in $1/\tau$ [147]:

$$W_{\mathrm{diss}} = -\beta \int_0^\tau dt\ \mathrm{tr}[\dot{\hat{P}}_t \mathscr{L}_t^+ \mathbb{J}_{\hat{\pi}_t}(\dot{\hat{P}}_t)] + \mathcal{O}(1/\tau^2), \tag{4.51}$$

For practical purposes, we need a way of computing the Drazin inverse. A useful representation of the Drazin inverse is provided by the following lemma:

Lemma 4.3.1. *Let $\mathscr{L}_t^+$ satisfy conditions (i)-(iii). Then the Drazin inverse admits the following unique integral representation:*

$$\mathscr{L}_t^+[\hat{A}] := \int_0^\infty d\nu\ e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\hat{A}] - \hat{A}\big). \tag{4.52}$$

Proof. We introduce the following trial solution

$$\Lambda_t^+[\hat{A}] := \int_0^\infty d\nu \; e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\hat{A}] - \hat{A}\big). \tag{4.53}$$

We first check (ii), which gives

$$\Lambda_t^+[\hat{\pi}_t] = \int_0^\infty d\nu \; e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\hat{\pi}_t] - \hat{\pi}_t\big) = 0, \tag{4.54}$$

which follows from the normalisation of $\hat{\pi}_t$. For (iii) we find

$$\begin{aligned} \mathrm{tr}[\Lambda_t^+[\hat{A}]] &= \int_0^\infty d\nu \; \mathrm{tr}[e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\hat{A}] - \hat{A}\big)], \\ &= \int_0^\infty d\nu \; \big(\mathrm{tr}[\hat{\pi}_t]\mathrm{tr}[\hat{A}] - \mathrm{tr}[\hat{A}]\big) = 0, \end{aligned} \tag{4.55}$$

where we used the fact that the propagator $e^{\nu\mathscr{L}_t}$ is trace-preserving. Regarding (i), we first find the following:

$$\begin{aligned} \mathscr{L}_t\Lambda_t^+[\hat{A}] &= \int_0^\infty d\nu \; \frac{d}{d\nu} e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\hat{A}] - \hat{A}\big), \\ &= \int_{\nu=0}^{\nu=\infty} d\big(e^{\nu\mathscr{L}_t}\big)\big(\hat{\pi}_t \mathrm{tr}[\hat{A}] - \hat{A}\big), \\ &= \hat{A} - \hat{\pi}_t \mathrm{tr}[\hat{A}] + \lim_{\nu\to\infty} e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\hat{A}] - \hat{A}\big), \\ &= \hat{A} - \lim_{\nu\to\infty} e^{\nu\mathscr{L}_t}\hat{A}, \\ &= \hat{A} - \hat{\pi}_t \mathrm{tr}[\hat{A}], \end{aligned} \tag{4.56}$$

where we used the fact that

$$\lim_{\nu\to\infty} e^{\nu\mathscr{L}_t}[\hat{B}] = \hat{\pi}_t, \quad \forall t \tag{4.57}$$

for any normalised operator $\hat{B}$. Finally, it also follows that

$$\begin{aligned} \Lambda_t^+\mathscr{L}_t[\hat{A}] &= \int_0^\infty d\nu \; e^{\nu\mathscr{L}_t}\big(\hat{\pi}_t \mathrm{tr}[\mathscr{L}_t[\hat{A}]] - \mathscr{L}_t[\hat{A}]\big), \\ &= -\int_0^\infty d\nu \; e^{\nu\mathscr{L}_t}\mathscr{L}_t[\hat{A}], \\ &= -\int_{\nu=0}^{\nu=\infty} d\big(e^{\nu\mathscr{L}_t}\big)\hat{A}, \\ &= \hat{A} - \lim_{\nu\to\infty} e^{\nu\mathscr{L}_t}\hat{A}, \\ &= \hat{A} - \hat{\pi}_t \mathrm{tr}[\hat{A}]. \end{aligned} \tag{4.58}$$

We thus conclude that $\Lambda_t^+ = \mathscr{L}_t^+$. ■

The equation (4.51) demonstrates that in the quasi-static regime, the dissipated work is related to a particular form of quantum covariance in power, with an additional dependence on the Drazin inverse of the system dynamics. As expected, in the infinitely slow limit where the system remains in an instantaneous thermal state, the dissipated work vanishes

$$\lim_{\tau\to\infty} W_{\text{diss}} = 0, \tag{4.59}$$

which provides the condition under which the second law can be saturated.

In order to check the validity of the fluctuation-dissipation relation (4.1), we next consider the work fluctuations for slow driving, with the duration τ large but finite. To evaluate the work fluctuations one can combine Lemma 4.2.1 with the perturbative expansion (4.49), resulting in the following lemma:

Lemma 4.3.2. *Let $\hat{\rho}_t$ satisfy the Lindblad equation* (4.30) *with the stationarity condition* (4.41)*. Then the work fluctuations up to first order in $1/\tau$ are given by*

$$\sigma_w^2 = -2\int_0^\tau dt\ \mathrm{tr}[\hat{P}_t \mathscr{L}_t^+ \mathbb{S}_{\hat{\pi}_t}(\hat{P}_t)] + \mathcal{O}(1/\tau^2), \tag{4.60}$$

where we have defined the linear operator

$$\mathbb{S}_{\hat{\varrho}}(\hat{A}) := \frac{1}{2}\{\hat{\varrho}, \Delta_{\hat{\varrho}}\hat{A}\}. \tag{4.61}$$

Proof. Firstly, the squared average work (4.28) can be written as follows:

$$\begin{aligned}\langle w\rangle^2 &= \int_0^\tau dt \int_0^\tau dt'\ \mathrm{tr}[\hat{P}_{t'}\ \hat{\rho}_{t'}]\mathrm{tr}[\hat{P}_t\ \hat{\rho}_t],\\ &= \int_0^\tau dt \int_0^\tau dt'\ \mathrm{Re}\Big(\mathrm{tr}[\hat{P}_{t'}\overleftarrow{P}(t',t)[\hat{\rho}_t]]\ \mathrm{tr}[\hat{P}_t\hat{\rho}_t]\Big),\end{aligned} \tag{4.62}$$

Combining Lemma 4.2.1 and (4.62) yields

$$\begin{aligned}\sigma_w^2 &= \langle w^2\rangle - \langle w\rangle^2,\\ &= \int_0^\tau dt \int_0^\tau dt'\ \mathrm{Re}\ \Big(\mathrm{tr}[\hat{P}_{t'}\ \overleftarrow{P}(t',t)[\Delta_{\hat{\rho}_t}\hat{P}_t\ \hat{\rho}_t]]\Big),\\ &= \frac{1}{2}\int_0^\tau dt \int_0^\tau dt'\ \mathrm{tr}[\hat{P}_{t'}\ \overleftarrow{P}(t',t)[\{\Delta_{\hat{\rho}_t}\hat{P}_t, \hat{\rho}_t]\}],\\ &= \int_0^\tau dt \int_0^\tau dt'\ \mathrm{tr}[\hat{P}_{t'}\overleftarrow{\mathcal{P}}(t',t)\ \mathbb{S}_{\hat{\rho}_t}(\hat{P}_t)],\end{aligned} \tag{4.63}$$

Recalling the definition of $\overleftarrow{\mathcal{P}}(t',t) = \overleftarrow{\mathcal{T}}\exp\big(\int_t^{t'} d\nu \mathscr{L}_\nu\big)$, we see that the integrand in (4.63) will decay to zero exponentially fast in $|t'-t| \sim \tau$ due to $\mathbb{S}_{\hat{\rho}_t}(\hat{P}_t)$ being traceless. This

means we can replace at first order in $1/\tau$ the time-dependent Liouvillian with the initial one:

$$\begin{aligned}\sigma_w^2 &\simeq \int_0^\tau dt \int_0^\tau dt' \ \mathrm{tr}[\hat{P}_{t'} e^{(t'-t)\mathscr{L}_t} \ \mathbb{S}_{\hat{\rho}_t}(\hat{P}_t)], \\ &= 2\int_0^\tau dt' \int_0^{t'} ds \ \mathrm{tr}[\hat{P}_{t'} \ e^{s\mathscr{L}_{t'-s}} \ \mathbb{S}_{\hat{\rho}_{t'-s}}(\hat{P}_{t'-s})], \end{aligned} \tag{4.64}$$

where in the second line we made the substitution $s = t' - t$. Since s will be typically much bigger than the thermalisation timescales, we can replace $t' - s$ with t' in all expressions and send the limit of the integration to infinity, since the correction for finite s will be exponentially suppressed. Then we get the following:

$$\begin{aligned}\sigma_w^2 &\simeq 2\int_0^\tau dt' \int_0^\infty ds \ \mathrm{tr}[\hat{P}_{t'} \ e^{s\mathscr{L}_{t'}} \ \mathbb{S}_{\hat{\rho}_{t'}}(\hat{P}_{t'})], \\ &= -2\int_0^\tau dt \ \mathrm{tr}[\hat{P}_t \mathscr{L}_t^+ \mathbb{S}_{\hat{\rho}_t}(\hat{P}_t)], \end{aligned} \tag{4.65}$$

where in the last step we used Lemma 4.3.1 for the Drazin inverse. Finally, at first order in $1/\tau$, we can substitute $\hat{\rho}_t \simeq \hat{\pi}_t$ using (4.49). ■

Lemma 4.3.2 now implies that the fluctuations also decay to zero in the infinite time limit, so that

$$\lim_{\tau\to\infty} \sigma_w = 0. \tag{4.66}$$

This conclusively proves that the quasi-static work distribution approaches a delta function, namely

$$\lim_{\tau\to\infty} p(w) \propto \delta(w - \Delta F), \tag{4.67}$$

which is the expected behaviour from classical stocahstic thermodynamics. However, in the first order regime with $\tau^2 \gg 1$, we will observe significant deviations from classical thermodynamics. To see this I will present the quantum generalisation of the work FDR (4.1). The crucial observation to make is the fact that in general, $\hat{H}_t$ will not commute with $\hat{P}_t$, implying

$$[\hat{H}_t, \hat{P}_t] \neq 0 \Leftrightarrow \mathbb{S}_{\hat{\pi}_t}(\hat{P}_t) \neq \mathbb{J}_{\hat{\pi}_t}(\hat{P}_t). \tag{4.68}$$

This fact leads us to the quantum work fluctuation-dissipation relation:

Theorem 4.3.1. *Consider a slowly-driven quantum system obeying the Lindblad master equation* (4.30) *with thermal fixed point* (4.41)*. Neglecting terms of order* $1/\tau^2$*, with* τ *the total duration of the protocol, the dissipated work and fluctuations are related according to*

$$W_{diss} = \frac{1}{2}\beta\sigma_w^2 - \beta\int_0^\tau dt\ \mathcal{Q}_t(\hat{\pi}_t, \hat{P}_t), \tag{4.69}$$

where

$$\mathcal{Q}_t(\hat{\varrho}, \hat{A}) := \int_0^1 da\ \mathrm{tr}\big[\hat{A}\mathscr{L}_t^+\big(\mathbb{J}_{\hat{\varrho}}(\hat{A}) - \mathbb{S}_{\hat{\varrho}}(\hat{A})\big)\big], \tag{4.70}$$

is referred to as the dynamical skew information.

Proof. One simply combines Lemma 4.3.2 with the definitions of the linear operators $\mathbb{S}_{\hat{\varrho}}(\hat{A})$ and $\mathbb{J}_{\hat{\varrho}}(\hat{A})$. By using the cyclicity of the trace, straightforward algebra completes the derivation. ■

As the main finding of this chapter, we conclude that in the quantum regime, the work fluctuations are no longer proportional to the dissipation, in contrast to classical thermodynamics. This is due to the non-commutativity between power and energy, resulting in an additional quantum correction term $\mathcal{Q}_t(\hat{\pi}_t, \hat{P}_t)$ appearing in the FDR. Remarkably, this quantum correction is related to the average Wigner-Yanase-Dyson skew information (3.15). To see this, let us consider an open system connected to a perfectly thermalising bath, in which case the Lindbladian is given by

$$\mathscr{L}_t[\hat{\rho}_t] = \Gamma^{-1}(\hat{\pi}_t - \hat{\rho}_t), \tag{4.71}$$

with a single inverse time-scale Γ^{-1}. It can then be shown from Lemma 4.3.1 that the Drazin inverse becomes $\mathscr{L}_t^+ = \mathrm{tr}[(.)]\hat{\pi}_t - \Gamma\hat{\mathbb{I}}$. In this case (4.70) becomes proportional to the average Wigner-Yanase-Dyson skew information (3.15):

$$\mathcal{Q}_t(\hat{\pi}_t, \hat{P}_t) := -\frac{\Gamma}{2}\int_0^1 da\ \mathrm{tr}\big[[\hat{P}_t, \hat{\pi}_t^a][\hat{P}_t, \hat{\pi}_t^{1-a}]\big], \tag{4.72}$$

For more general dynamics, one may show that the functional $\mathcal{Q}_t(\hat{\varrho}, \hat{A})$ retains the same properties as the usual skew information. This is summarised in the following theorem:

Theorem 4.3.2. *Consider a system governed by the time-dependent Lindblad equation* (4.42) *that satisfies the properties of detailed balance. Then the dynamical skew information is positive semi-definite;* $\mathcal{Q}_t(\hat{\pi}_t, \hat{P}_t) \geq 0$ *and is given by*

$$\mathcal{Q}_t(\hat{\pi}_t, \hat{P}_t) := \tau_t^{eq} \times \mathcal{Q}(\hat{\pi}_t, \hat{P}_t), \tag{4.73}$$

with $\mathcal{Q}(\hat{\pi}_t, \hat{P}_t)$ the usual average WYD skew information and

$$\tau_t^{eq} := \int_0^\infty d\nu \, \frac{\langle\langle \hat{P}_t(\nu), \hat{P}_t(0)\rangle\rangle}{\langle\langle \hat{P}_t(0), \hat{P}_t(0)\rangle\rangle}, \tag{4.74}$$

is the quantum integral relaxation time. We denote $\langle\langle \hat{A}_t, \hat{B}_t\rangle\rangle = \mathrm{tr}[\hat{B}_t^\dagger(\mathbb{S}_{\hat{\pi}_t} - \mathbb{J}_{\hat{\pi}_t})(\hat{A}_t)]$ as the quantum correlation function and $\hat{A}_t(\nu) = e^{\nu\mathscr{L}_t^\dagger}(\hat{A}_t)$ is the evolution of $\hat{A}_t$ in the Heisenberg picture.

Proof. Consider the Hilbert space $\mathbb{M}^d$ of $d \times d$ complex matrices with Hilbert-Schmidt inner product $\langle \hat{A}, \hat{B}\rangle = \mathrm{tr}[\hat{B}^\dagger \hat{A}]$. Then any superoperator $\mathcal{M}(.)$ acting on the elements of this Hilbert space can be expressed as a $d^2 \times d^2$ matrix. The matrix describing $\mathcal{M}(.)$ is positive if $\mathrm{tr}[\hat{A}^\dagger \mathcal{M}(\hat{A})] \geq 0$ for any $\hat{A} \in \mathbb{M}^d$, and we define the adjoint $\mathcal{M}^\dagger(.)$ as the superoperator satisfying $\mathrm{tr}[\mathcal{M}^\dagger(\hat{B}^\dagger)\hat{A}] = \mathrm{tr}[\hat{B}^\dagger \mathcal{M}(\hat{A})]$.

Now note that the dynamical skew information is a real-valued trace functional, thus it is sufficient to prove positivity of the quantity

$$\mathscr{I}(\hat{\pi}_t, \hat{A}) := -\mathrm{Re}\ \mathrm{tr}[\hat{A}\mathscr{L}_t^+ \, \mathcal{M}_t(\hat{A})], \tag{4.75}$$

where $\hat{A} = \hat{A}^\dagger$ is an arbitrary hermitian operator, $\mathscr{L}_t^+$ the Drazin inverse of the Lindbladian defined in (4.52) and

$$\mathcal{M}_t(.) := \frac{1}{2}\{\hat{\pi}_t, (.)\} - \int_0^1 ds \ \hat{\pi}_t^s(.)\hat{\pi}_t^{1-s}. \tag{4.76}$$

Here $\mathcal{M}_t(.)$ represents the difference between the arithmetic and logarithmic matrix means, and is hence a positive superoperator due to the Kubo-Ando inequality [162].

Since $\hat{\pi}_t$ commutes with Hamiltonian $\hat{H}_t$ one can verify the commutation relation

$$[\mathcal{M}_t(.), \theta_t(.)] = 0. \tag{4.77}$$

Let us now consider the relation between $\mathcal{M}_t(.)$ and the dissipator $\mathscr{D}_t(.)$. It is first useful to introduce the following integral representation for the matrix power $\hat{\pi}^s$ for positive $\hat{\pi}$ [163]:

$$\hat{\pi}^s = \int_0^\infty d\mu_s(x)\big(e^{-x\hat{\pi}} - \hat{\mathbb{I}}\big); \quad s \in (0,1). \tag{4.78}$$

with $\mu_s(x)$ a positive measure on $(0,\infty)$ that we leave unspecified for convenience. Using detailed balance (4.46) we get the following:

$$\begin{aligned}\hat{\pi}_t^s \hat{A}_\alpha(\omega_t) &= \int_0^\infty d\mu_s(x)\big(e^{-x\hat{\pi}_t} - \hat{\mathbb{I}}\big)\hat{A}_\alpha(\omega_t),\\ &= \int_0^\infty d\mu_s(x)\bigg(\sum_{n=0}^\infty \frac{(-x)^n\hat{\pi}_t^n)}{n!} - \hat{\mathbb{I}}\bigg)\hat{A}_\alpha(\omega_t),\\ &= \hat{A}_\alpha(\omega_t)\int_0^\infty d\mu_s(x)\bigg(\sum_{n=0}^\infty \frac{(-x)^n(e^{\beta\omega_t}\hat{\pi}_t)^n)}{n!} - \hat{\mathbb{I}}\bigg),\\ &= e^{s\beta\omega_t}\hat{A}_\alpha(\omega_t)\hat{\pi}_t^s. \qquad (4.79)\end{aligned}$$

Similarly one finds

$$\hat{\pi}_t^s \hat{A}_\alpha^\dagger(\omega_t) = e^{-s\beta\omega_t}\hat{A}_\alpha^\dagger(\omega_t)\hat{\pi}_t^s. \qquad (4.80)$$

This then implies

$$\hat{A}_{\alpha'}(\omega_t)\mathcal{M}_t(.)\hat{A}_\alpha^\dagger(\omega_t) = e^{-\beta\omega_t}\mathcal{M}_t\hat{A}_{\alpha'}(\omega_t)(.)\hat{A}_\alpha^\dagger(\omega_t) \qquad (4.81)$$

Using this one obtains the following:

$$\begin{aligned}\mathscr{D}_t\mathcal{M}_t(.) &= \sum_{\omega_t}\sum_{\alpha,\alpha'}\gamma_{\alpha\alpha'}(\omega_t)\bigg(\hat{A}_{\alpha'}(\omega_t)\mathcal{M}_t(.)\hat{A}_\alpha^\dagger(\omega_t) - \frac{1}{2}\{\hat{A}_\alpha^\dagger(\omega_t)\hat{A}_{\alpha'}(\omega_t),\mathcal{M}_t(.)\}\bigg),\\ &= \mathcal{M}_t\sum_{\omega_t}\sum_{\alpha,\alpha'}\gamma_{\alpha\alpha'}(\omega_t)e^{-\beta\omega_t}\hat{A}_{\alpha'}(\omega_t)(.)\hat{A}_\alpha^\dagger(\omega_t) - \frac{1}{2}\mathcal{M}_t\sum_{\omega_t}\sum_{\alpha,\beta}\gamma_{\alpha\alpha'}(\omega_t)\{\hat{A}_\alpha^\dagger(\omega_t)\hat{A}_{\alpha'}(\omega_t),(.)\},\\ &= \mathcal{M}_t\sum_{\omega_t}\sum_{\alpha,\alpha'}\gamma_{\alpha'\alpha}(-\omega_t)\hat{A}_{\alpha'}^\dagger(-\omega_t)(.)\hat{A}_\alpha(-\omega_t) - \frac{1}{2}\mathcal{M}_t\sum_{\omega_t}\sum_{\alpha,\alpha'}\gamma_{\alpha\alpha'}(\omega_t)\{\hat{A}_\alpha^\dagger(\omega_t)\hat{A}_{\alpha'}(\omega_t),(.)\},\\ &= \mathcal{M}_t\sum_{\omega_t}\sum_{\alpha,\alpha'}\gamma_{\alpha\alpha'}(\omega_t)\hat{A}_\alpha^\dagger(\omega_t)(.)\hat{A}_{\alpha'}(\omega_t) - \frac{1}{2}\mathcal{M}_t\sum_{\omega_t}\sum_{\alpha,\alpha'}\gamma_{\alpha\alpha'}(\omega_t)\{\hat{A}_\alpha^\dagger(\omega_t)\hat{A}_{\alpha'}(\omega_t),(.)\},\\ &= \mathcal{M}_t\mathscr{D}_t^\dagger(.),\end{aligned}$$

where in the second line we used (4.81), in the third line we used (4.44) and (4.45), in the fourth line swapped indices $-\omega_t \to \omega_t$ and $\alpha \to \alpha'$, and in the final line used the definition of the adjoint superoperator $\mathrm{tr}[\mathscr{D}_t^\dagger(\hat{B}^\dagger)\hat{A}] = \mathrm{tr}[\hat{B}^\dagger\mathscr{D}_t(\hat{A})]$ and the fact that the second term is self-adjoint.

We next introduce a complimentary Lindbladian of form

$$\tilde{\mathscr{L}}_t(.) = \theta_t(.) + \mathscr{D}_t^\dagger(.). \qquad (4.82)$$

By using (4.77) and the commutation relation for the dissipator we have

$$\mathscr{L}_t\mathcal{M}_t(\hat{A}) = \mathcal{M}_t\tilde{\mathscr{L}}_t(\hat{A}). \tag{4.83}$$

Notably the real part of the spectrum of $\tilde{\mathscr{L}}_t(.)$ coincides with that of the original Lindbladian $\mathcal{L}_t(.)$ due to the fact that $\theta_t(.)$ is skew hermitian. Therefore both $\tilde{\mathscr{L}}_t(.)$ and $\mathscr{L}_t(.)$ share a unique zero eigenvalue associated with their shared stationary state $\hat{\pi}_t$. Therefore we can define a pair of Drazin inverses using Lemma 4.3.1:

$$\mathscr{L}_t^+(\hat{A}) = -\int_0^\infty d\nu\ e^{\nu\mathscr{L}_t}(\hat{A} - \mathrm{tr}[\hat{A}]\hat{\pi}_t), \tag{4.84}$$

$$\tilde{\mathscr{L}}_t^+(\hat{A}) = -\int_0^\infty d\nu\ e^{\nu\tilde{\mathscr{L}}_t}(\hat{A} - \mathrm{tr}[\hat{A}]\hat{\pi}_t). \tag{4.85}$$

for any $\hat{A} \in \mathbb{M}^d$. These inverses act according to

$$\begin{aligned}\mathscr{L}_t\mathscr{L}_t^+(\hat{A}) &= \mathscr{L}_t^+\mathscr{L}_t(\hat{A}) = \hat{A} - \hat{\pi}_t\mathrm{tr}[\hat{A}],\\ \tilde{\mathscr{L}}_t\tilde{\mathscr{L}}_t^+(\hat{A}) &= \tilde{\mathscr{L}}_t^+\tilde{\mathscr{L}}_t(\hat{A}) = \hat{A} - \hat{\pi}_t\mathrm{tr}[\hat{A}].\end{aligned} \tag{4.86}$$

For any traceless matrix $\{\hat{B} \mid \hat{B} \in \mathbb{M}^d,\ \mathrm{tr}[\hat{B}] = 0\}$, we can combine (4.86) and (4.83) to get

$$\mathscr{L}_t^+\mathcal{M}_t(\hat{B}) = \mathcal{M}_t\tilde{\mathscr{L}}_t^+(\hat{B}), \tag{4.87}$$

We also define the following superoperator:

$$\mathscr{V}_t(.) := -\frac{\mathscr{L}_t^+ + [\tilde{\mathscr{L}}_t^+]^\dagger}{2}(.). \tag{4.88}$$

Using (4.87) one can also see that

$$\mathscr{V}_t\mathcal{M}_t(\hat{B}) = \mathcal{M}_t\mathscr{V}_t^\dagger(\hat{B}). \tag{4.89}$$

Furthermore, since by assumption the real part of the eigenvalues of the Lindbladian $\mathscr{L}_t$ are negative, the same holds true for both Drazin inverses $\mathscr{L}_t^+$ and $\tilde{\mathscr{L}}_t^+$. As a result, $\mathscr{V}_t(.)$ contains no eigenvalues with a negative real part. It is known in linear algebra that a matrix product XY with $Y \geq 0$ is positive if the eigenvalues of X have no negative real part and $XY = YX^\dagger$. Since $\mathcal{M}_t(.)$ is positive with respect to the Hilbert-Schmidt scalar product, (4.89) implies that

$$\mathscr{V}_t\mathcal{M}_t \geq 0. \tag{4.90}$$

Finally, we return to the trace functional (4.75). Let us introduce the projection onto the traceless subspace $\mathcal{P}_T(\hat{A}) = \hat{A} - \mathrm{tr}[\hat{A}]\hat{\mathbb{I}}/d$. Taking all results together one gets

$$\begin{aligned}
\mathscr{I}(\hat{\pi}_t, \hat{A}) &= -\mathrm{Re}\ \mathrm{tr}[\hat{A}\mathscr{L}_t^{+}\mathcal{M}_t(\hat{A})], \\
&= -\mathrm{Re}\ \mathrm{tr}[\mathcal{P}_T(\hat{A})\mathscr{L}_t^{+}\mathcal{M}_t\mathcal{P}_T(\hat{A})], \\
&= -\frac{1}{2}\mathrm{tr}[\hat{B}[\mathscr{L}_t^{+}\mathcal{M}_t + \mathcal{M}_t[\mathscr{L}_t^{+}]^{\dagger}](\hat{B})], \\
&= -\frac{1}{2}\mathrm{tr}[\hat{B}[\mathscr{L}_t^{+} + [\tilde{\mathscr{L}}_t^{+}]^{\dagger}]\mathcal{M}_t(\hat{B})], \\
&= \mathrm{tr}[\hat{B}\mathcal{V}_t\mathcal{M}(\hat{B})], \\
&\geq 0, \qquad (4.91)
\end{aligned}$$

where in the second line we used the fact that only traceless elements contribute to the functional, in the third line we set $\mathcal{P}_T(\hat{A}) = \hat{B}$ and $\mathcal{M}_t^{\dagger}(.) = \mathcal{M}_t(.)$, in the fourth line we used (4.87) and in the final line we used the matrix positivity (4.90). Since the above holds for any hermitian matrix $\hat{A}$, we conclude that the dynamical skew information (4.70) is positive. To arrive at the final form of the skew information we simply substitute in the integral expression for the Drazin inverse.

■

Theorem 4.3.2 demonstrates that the dynamical skew information is simply given by the usual skew information weighted by an additional relaxation time. This integral relaxation time τ_t^{eq} simply measures the area under the curve for the dynamical quantum correlation function $\langle\langle \hat{P}_t(\nu), \hat{P}_t(0)\rangle\rangle$ in and away from equilibrium. More simply, τ_t^{eq} measures the effective timescale over which the quantum power fluctuations decay to their equilibrium values. This should be compared with the usual form of of the integral relaxation time found in classical systems [148]. To summarise, the quantum counterpart to the work fluctuation-dissipation (4.1) can be expressed as

$$W_{\mathrm{diss}} = \frac{1}{2}\beta\sigma_w^2 - \Phi_w, \qquad (4.92)$$

with our quantum correction given by

$$\Phi_w = \beta \int_0^{\tau} dt\ \tau_t^{eq} \times \mathcal{Q}(\hat{\pi}_t, \hat{P}_t). \qquad (4.93)$$

As a consequence of the positivity of Φ_w we conclude that in the quantum regime the work fluctuations are actually greater than the dissipation, namely

$$\sigma_w^2 \geq 2k_B T\ W_{\mathrm{diss}}. \qquad (4.94)$$

These excess fluctuations are clearly of a quantum nature, and they can even be witnessed directly from the work distribution itself. While we saw in the classical scenario a quasi-static process will result in a Gaussian work distribution $p(w) \propto \exp[(w - \langle w \rangle)^2/2\sigma_w^2]$ [54, 55], the quantum work FDR in Theorem 4.3.1 implies that this no longer holds true for quantum systems. To prove this we can use the Jarzynski equality (4.23), which relates the change in equilibrium free energy to the cumulants of work done on the system:

$$\Delta F = -\beta^{-1} \ln\langle e^{-\beta w} \rangle = \sum_{k=1}^{\infty} \frac{(-\beta)^{k-1}}{k!} \kappa_w^k. \tag{4.95}$$

Here κ_w^k represent the work cumulants with $\kappa_w^1 = \langle w \rangle$ and $\kappa_w^2 = \sigma_w^2$. After rearranging terms in (4.95) and combining this with (4.92), we find

$$\sum_{k=3}^{\infty} \frac{(-\beta)^k}{k!} \kappa_W^k = \Phi_w \geq 0, \tag{4.96}$$

where the inequality follows from the positivity of the dynamical skew information. Due to the properties of the skew information, the RHS of (4.96) vanishes *iff* $[\hat{P}_t, \hat{H}_t] = 0 \ \ \forall t \in [0, \tau]$. Since a Gaussian work distribution has zero cumulants for $k \geq 3$, we conclude that $p(w)$ is necessarily non-Gaussian whenever the process generates coherences in power. Remarkably, (4.96) suggests that the work distribution provide a direct means of observing quantum fluctuations in power in the quasi-static regime, since any deviation from a Gaussian shape is equivalent to a witness of quantum coherence.

4.4 Geometry of quantum work fluctuations and optimal paths

We have now derived an expression for work fluctuations in Lemma 4.3.2 for slow dynamics, and this has been shown to differ from the equivalent expression for dissipation (4.51). In analogy to classical thermodynamic geometry summarised by the formula (4.9), one can now seek to express our results in terms of metrics over the manifold of quantum thermal states. To do this we parameterise our system Hamiltonian in terms of a set of conjugate forces, so that $\hat{H}_t = \vec{\lambda}_t \cdot \vec{X}$, with $\vec{\lambda}_t = (\lambda_1(t), \lambda_2(t), ...)$ a vector of scalar control parameters and $\vec{X} = (\hat{X}_1, \hat{X}_2, ...)$ the forces, which generally will not commute. In addition, it will be helpful to apply a simple change of variables so that $\vec{\lambda}_t \to \vec{\lambda}_{t\tau}$ and re-parameterise the Lindbladian $\mathscr{L}_\lambda$ and thermal state π_λ in terms of the control

parameters. Applying this to the dissipation (4.51), one obtains the following:

$$W_{\text{diss}} = \frac{1}{\tau}\int_0^1 dt\ \left[\frac{d\vec{\lambda}}{dt}\right]^T .\xi(\vec{\lambda}).\left[\frac{d\vec{\lambda}}{dt}\right], \tag{4.97}$$

where

$$\xi_{ij}(\vec{\lambda}) := -\frac{\beta}{2}\text{tr}[\hat{X}_i\mathscr{L}_\lambda^+\mathbb{J}_{\hat{\pi}_\lambda}(\hat{X}_j)] - \frac{\beta}{2}\text{tr}[\hat{X}_j\mathscr{L}_\lambda^+\mathbb{J}_{\hat{\pi}_\lambda}(\hat{X}_i)]. \tag{4.98}$$

If follows from the results in [147] that $\xi(\vec{\lambda})$ is a positive, symmetric metric tensor. This follows from the fact that the rate of dissipated work is equal to the entropy production rate up to first order in $1/\tau$, which is positive due to the contractivity of the quantum relative entropy [164]. Similarly for the fluctuations one finds

$$\sigma_w^2 = \frac{2k_BT}{\tau}\int_0^1 dt\ \left[\frac{d\vec{\lambda}}{dt}\right]^T .\Lambda(\vec{\lambda}).\left[\frac{d\vec{\lambda}}{dt}\right], \tag{4.99}$$

where

$$\Lambda_{ij}(\vec{\lambda}) := -\frac{\beta}{2}\text{tr}[\hat{X}_i\mathscr{L}_\lambda^+\mathbb{S}_{\hat{\pi}_\lambda}(\hat{X}_j)] - \frac{\beta}{2}\text{tr}[\hat{X}_j\mathscr{L}_\lambda^+\mathbb{S}_{\hat{\pi}_\lambda}(\hat{X}_i)]. \tag{4.100}$$

Using Theorem 4.3.2 we can also conclude that $\Lambda(\vec{\lambda})$ is a positive and symmetric metric tensor. This reveals a non-trivial quantum extension to the notion of thermodynamic geometry. If one considers the Riemannian manifold of quantum thermal states $\mathcal{M}_\lambda = \{\hat{\pi}_\lambda\}$, then these results reveal that the work fluctuations and dissipation are related to different metrics over the manifold. It is only in the classical scenario where $[\hat{X}_i, \hat{X}_j] = 0\ \forall i,j$ that both metrics reduce to the usual Fisher-Rao metric given in (4.9). We can further introduce two different definitions of quantum thermodynamic length, namely the quantities

$$\ell_\xi(\vec{\lambda}_0, \vec{\lambda_1}) = \int_0^1 dt\ \sqrt{\left[\frac{d\vec{\lambda}}{dt}\right]^T .\xi(\vec{\lambda}).\left[\frac{d\vec{\lambda}}{dt}\right]}, \tag{4.101}$$

for dissipation and

$$\ell_\Lambda(\vec{\lambda}_0, \vec{\lambda_1}) = \int_0^1 dt\ \sqrt{\left[\frac{d\vec{\lambda}}{dt}\right]^T .\Lambda(\vec{\lambda}).\left[\frac{d\vec{\lambda}}{dt}\right]}, \tag{4.102}$$

for the fluctuations. Both quantities represent distance measures between states in the manifold $\mathcal{M}_\lambda$ with geodesics satisfying the triangle inequality. It then follows from the

Cauchy-Schwarz inequality that the dissipation and fluctuations are bounded geometrically by these lengths according to

$$W_{\text{diss}} \geq \frac{\ell_\xi^2}{\tau}, \qquad \sigma_w^2 \geq \frac{2k_B T \ell_\Lambda^2}{\tau}. \tag{4.103}$$

Since both the dissipation (4.97) and work fluctuations (4.99) are related to metrics over the thermodynamic state space, one can seek find optimal finite-time protocols by minimising these quantities over the set of controllable Hamiltonian parameters. Much like in the classical scenario, one can view (4.97) and (4.99) as action functionals for curves $\vec{\lambda}_t$. Minimising quantum work fluctuations and dissipation is then tantamount to minimising the corresponding action using the Euler-Lagrange equations. Since the path of least action will also be the shortest, the optimal protocol $\vec{\lambda}_t^{opt}$ will move along the geodesics associated to the lengths (4.102) or (4.101). Such a process will saturate either of the inequalities (4.103), depending on whether one minimises W_{diss} or σ_w. Crucially, since the two metrics differ in general, these geodesics will no longer coincide in contrast to classical thermodynamics.

To see this, consider a single parameter change $\lambda_0 \to \lambda_1$ and $H_\lambda = \lambda\hat{X} + \hat{X}_0$, with $[\hat{X}_0, \hat{X}] \neq 0$. Denote the rescaled work fluctuations by $\tilde{\sigma}_w^2 = \frac{1}{2}\beta\sigma_w^2$. We wish to minimise the linear objective function

$$\mathcal{I}_\alpha := \alpha\tilde{\sigma}_w^2 + (1-\alpha)W_{\text{diss}}; \quad \alpha \in [0,1], \tag{4.104}$$

with respect to λ_t. Using the quantum FDR (4.92) we have

$$\mathcal{I}_\alpha = \frac{1}{\tau}\int_0^1 dt\ \dot{\lambda}_t^2\big(\xi(\lambda) + \alpha\beta\mathcal{Q}_\lambda(\hat{\pi}_\lambda, \hat{X})\big), \tag{4.105}$$

The functional (4.105) is minimised by the solution to the Euler-Lagrange equation:

$$\frac{\partial\mathcal{C}_\alpha}{\partial\lambda} = \frac{d}{dt}\left[\frac{\partial\mathcal{C}_\alpha}{\partial\dot{\lambda}}\right] \implies \ddot{\lambda} = -\frac{\mathcal{C}'_\alpha}{2\mathcal{C}_\alpha}\dot{\lambda}^2, \tag{4.106}$$

with cost function $\mathcal{C}_\alpha(\lambda,\dot{\lambda}) := \dot{\lambda}^2\big(\xi(\lambda)+\alpha\beta\mathcal{Q}_\lambda(\hat{\pi}_\lambda,\hat{X})\big)$. Solving (4.106) yields an equation for the optimal velocity of the control parameter for a given α:

$$\dot{\lambda}_t^{\text{opt}}(\alpha) = \frac{(\lambda_1-\lambda_0)\big(\xi(\lambda_t)+\alpha\beta\mathcal{Q}_\lambda(\hat{\pi}_t,\hat{X})\big)^{-1/2}}{\int_0^1 dt\ \big(\xi(\lambda_t)+\alpha\beta\mathcal{Q}_\lambda(\hat{\pi}_t,\hat{X})\big)^{-1/2}}, \tag{4.107}$$

One concludes that the optimal velocity is proportional to the following:

$$\dot{\lambda}_t^{\text{opt}}(\alpha) \propto \big(\xi(\lambda_t)+\alpha\beta\mathcal{Q}_\lambda(\hat{\pi}_\lambda,\hat{X})\big)^{-1/2}. \tag{4.108}$$

This means that during the optimal protocol, the combination of the classical power fluctuations (ie. $\dot{\lambda}_t^2\xi(\lambda_t)$) plus an additional contribution from the quantum power fluctuations need to be kept constant. This contrasts with the classical case where the optimal protocol only requires constant total fluctuations in power for any value of α [142].

The geometric expressions for dissipated work and fluctuations provide the central tools for computing optimal thermodynamic processes. To highlight the computational power of these results we will now consider a simple example with Lindbladian $\mathscr{L}_t[\hat{\rho}_t] = \Gamma^{-1}(\hat{\pi}_t - \hat{\rho}_t)$. Let us assume the driven system consists of a harmonic oscillator with a time-dependent frequency,

$$\hat{H}_{\omega_t} = \omega_t\left(\hat{a}^\dagger_{\omega_t}\hat{a}_{\omega_t} + \frac{1}{2}\right), \tag{4.109}$$

The resulting power operator is given by

$$\hat{P}_{\omega_t} = \dot{\omega}_t\left(\frac{\hat{H}_t}{\omega_t} + \frac{(\hat{a}^\dagger_{\omega_t})^2 + \hat{a}^2_{\omega_t}}{2}\right), \tag{4.110}$$

Clearly $[\hat{P}_{\omega_t}, \hat{H}_{\omega_t}] \neq 0$ and hence one should observe deviations from the classical work fluctuation-dissipation relation, as predicted by Theorem 4.3.1. To see this, we can straightforwardly compute the two metrics in (4.97) and (4.99) as

$$\xi(\omega_t) = \frac{\beta\omega_t + \sinh(\beta\omega_t)}{2\omega_t\cosh(\beta\omega_t) - 2\omega_t}\Gamma, \tag{4.111}$$

$$\Lambda(\omega_t) = \frac{\beta}{2}\coth^2\big(\frac{\beta\omega_t}{2}\big)\Gamma. \tag{4.112}$$

Figure 4.1 shows these quantities plotted as a function of β. It can be seen that in the high temperature limit ($\beta \to 0$), one has $\xi(\omega) \simeq \Lambda(\omega)$, with the skew information vanishing. This is expected considering that the high temperature limit can be considered equivalent to the classical limit, and the state of the system will approach the maximally mixed state which nullifies any quantum fluctuations in power. In the low temperature limit ($\beta \to \infty$) we see the dissipation metric $\xi(\omega)$ vanishes to zero due to the fact that the system remains in the ground state at all times. Secondly, one also observes that the fluctuations metric $\Lambda(\omega)$ converges to the skew information $\Lambda(\omega) - \xi(\omega)$. This is also expected considering that the skew information reduces to the variance in power for a pure state.

In Figure 4.2 I demonstrate the work fluctuation-dissipation relation given by Theorem 4.3.1 for a simple linear protocol

$$\omega_t = \omega_0 + t(\omega_\tau - \omega_0), \tag{4.113}$$

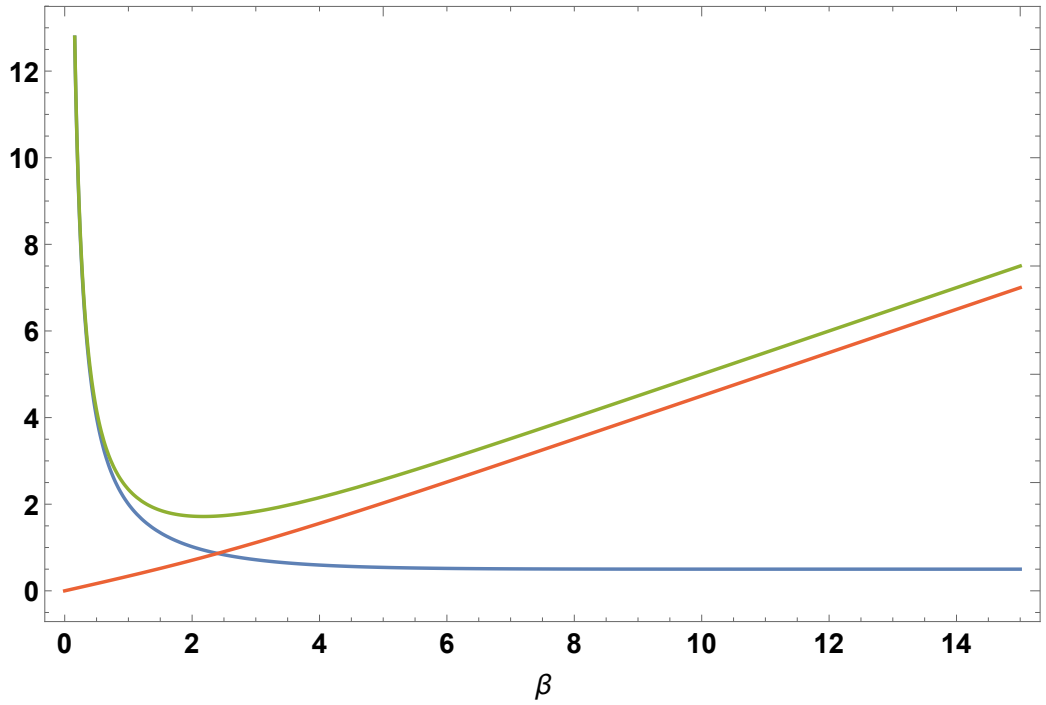


Figure 4.1: **Thermodynamic metrics for harmonic oscillator.** Plot of the different metrics given by (4.111) with $\Lambda(\omega)$ (green), $\xi(\omega)$ (blue) and the difference $\Lambda(\omega) - \xi(\omega)$ (red) as a function of inverse temperature and $\omega = 1$, $\Gamma = 1$.

plotted as a function of inverse temperature β. The area between the curves represents the quantum correction Φ_w appearing in (4.92), which is positive semi-definite. The correction is equivalent to the time-integrated skew information, and quantifies the quantum power fluctuations along the protocol. Intuitively the quantum term Φ_w increases monotonically from zero with increasing β, implying that quantum power fluctuations and violations of the classical FDR (4.1) dominate in the low temperature regime.

We next consider computing the optimal change in frequency $\omega_t^{opt}(\alpha)$ in order to minimise the fluctuations-dissipation trade-off function $\mathcal{I}_\alpha$ in (4.104). The general single-parameter solution has been calculated from the Euler-Lagrange in (4.92). In Figure 4.3 the geodesic curves $\omega_t^{opt}(\alpha)$ are plotted in units of β as a function of time. One clearly sees that the optimal protocol differs depending on α, demonstrating that the process of minimal dissipation does not coincide with the process of minimal fluctuations. Moreover, in Figure 4.4 the Pareto fronts for the linear objective function $\mathcal{I}_\alpha$ in (4.105) are shown for different temperatures. One sees that as temperature decreases, the curves grow wider indicating the increased relevance of quantum fluctuations and violations of the classical work FDR (4.1). At high enough temperatures the Pareto fronts converge to a single

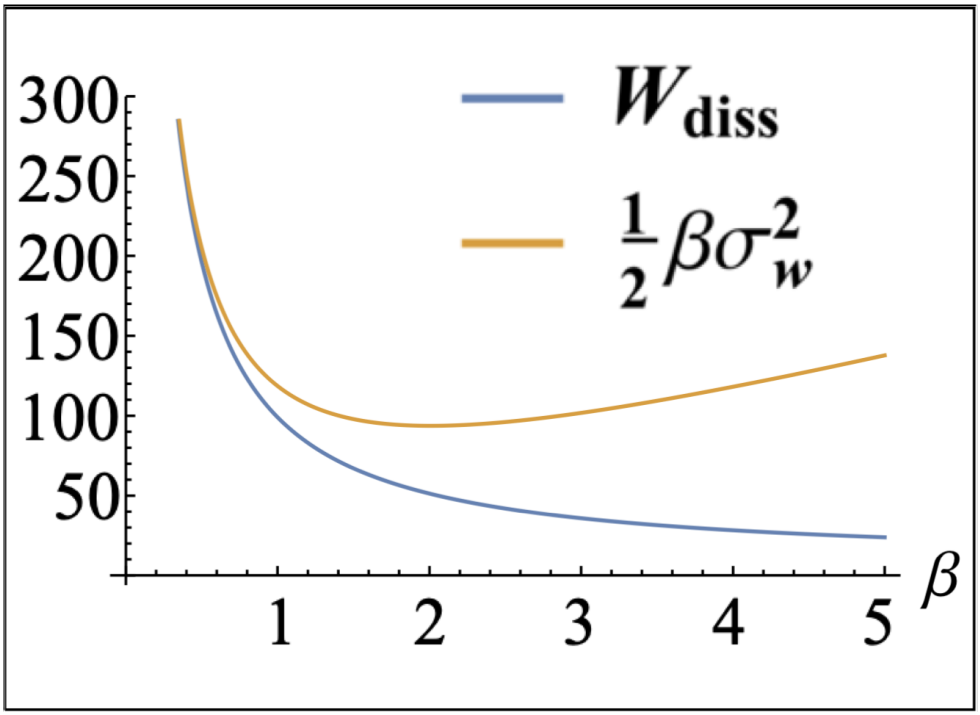


Figure 4.2: **Work FDR for harmonic oscillator.** Plot of dissipated work (blue) and work fluctuations (orange) for a linear change in the oscillator frequency. The difference between the curves represents the quantum term Φ_w appearing in (4.92).(Graph created by Matteo Scandi)

point. In this case the optimal values of W_{diss} and $\beta\sigma_w^2/2$ become equal as expected in the classical limit.

4.5 Concluding remarks

In this chapter I have investigated the effects of quantum mechanics on the work fluctuation-dissipation relation (4.1). In the classical regime this relation demonstrates that the dissipated work is proportional to the resulting work fluctuations when a system is driven slowly in time. Fundamentally this means that any optimal process with minimal dissipation will also minimise the unwanted fluctuations. Optimisation can be achieved through use of differential geometry, since the dissipated work rate can be related to the Fisher-Rao metric on the manifold of thermal states, as exemplified in (4.9). Physically, this connection between work dissipation and fluctuations stems from the fact that the work distribution becomes Gaussian in the slow driving limit, as proven in [55]. However, it had not previously been shown whether or not such a relation holds in the quantum

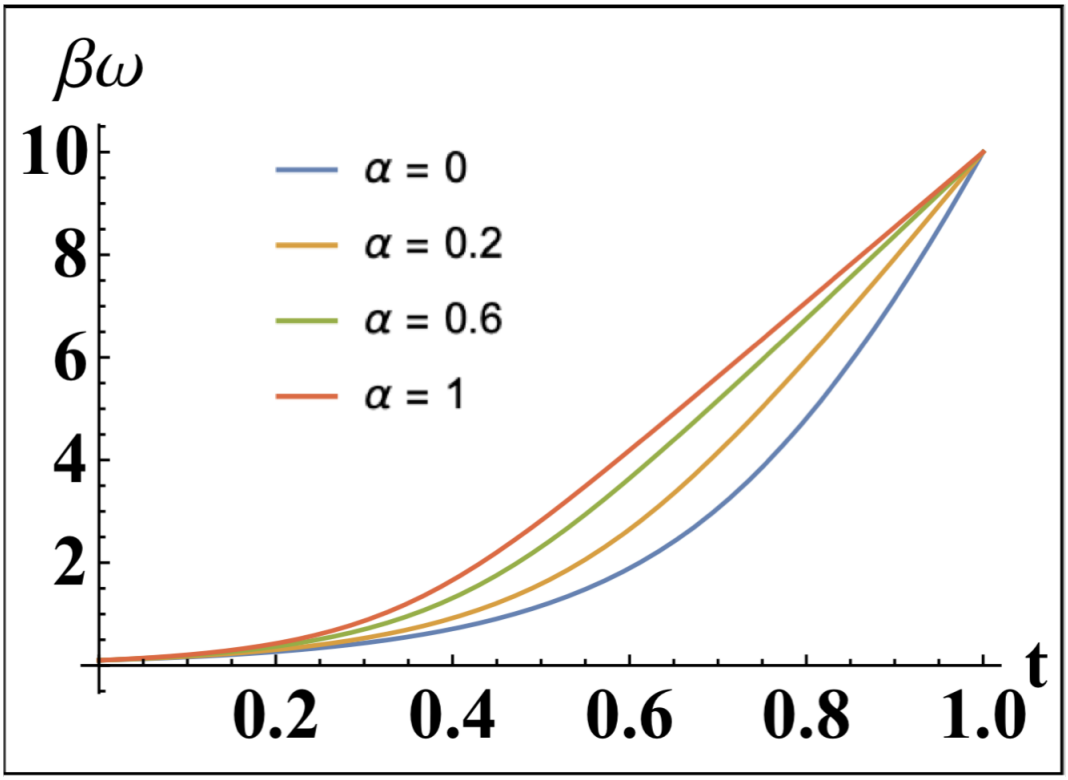


Figure 4.3: **Optimal work geodesics.** Plot of the optimal curves $\omega_t^{opt}(\alpha)$ (4.107) for the driven harmonic oscillator as a function of time for $\beta = 1$. The curves interpolate between the curve of minimal dissipation ($\alpha = 0$) and minimal fluctuations ($\alpha = 1$).(Graph created by Matteo Scandi)

setting, and this question was addressed in this chapter.

In quantum mechanical systems phase space trajectories are no longer well defined, and the standard definition of stochastic work as integrated power is modified. Instead, for a quantum system evolving in contact with an environment, work fluctuations are identified from probabilistic jumps between energy eigenstates due to quantum measurements. The so-called two-projective measurement scheme leads to a consistent thermodynamic framework in which the second law and fluctuation relations can be derived [143]. Furthermore, this definition shares a correspondence with stochastic thermodynamics in the semi-classical limit [155]. In Section 4.2 I reviewed this theoretical framework for defining quantum work and derived explicit expressions for the average work and its fluctuations in a generic open system evolving under Lindbladian dynamics, as shown in Lemma 4.2.1. It was shown that, while in general obtaining the statistics of work in an open system requires full knowledge of the unitary dynamics, under the Born-Markov assumption one may obtain the first and second cumulants of work from the reduced

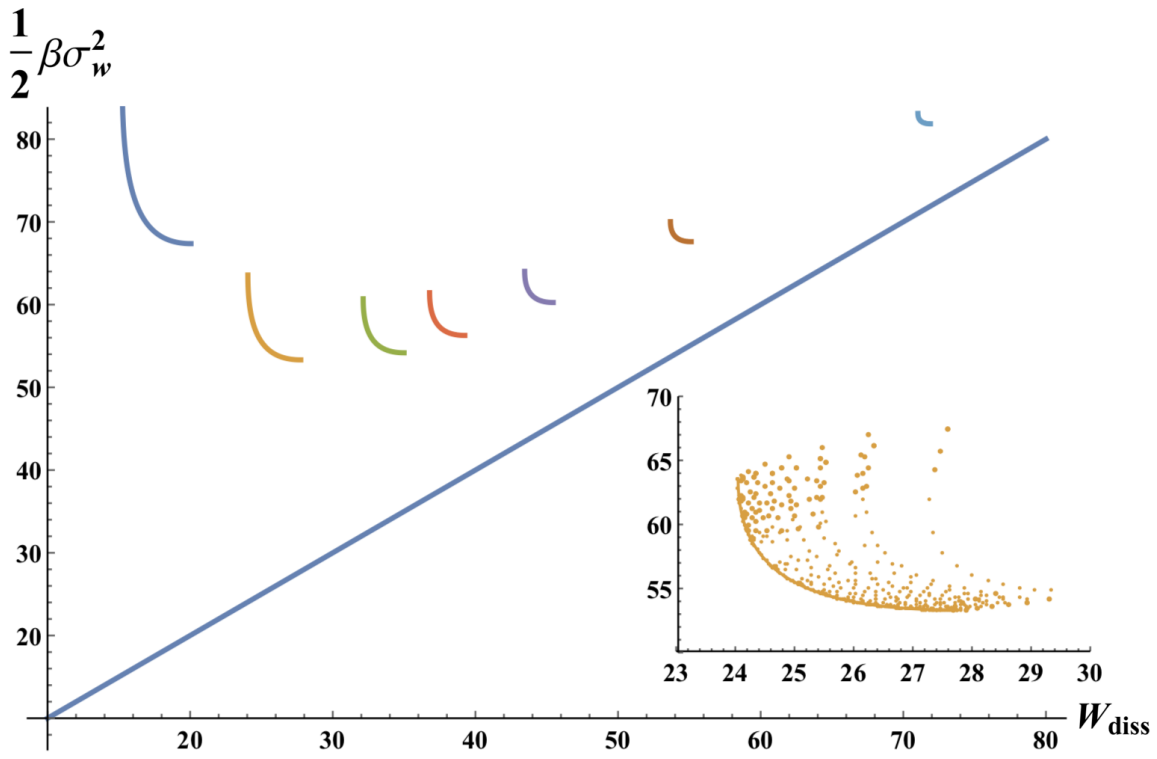


Figure 4.4: **Pareto fronts for the optimal processes.** Plot of Pareto fronts for the minimal values of $\mathcal{I}_\alpha$ in (4.105) over all $\alpha \in [0, 1]$ for the harmonic oscillator at different temperatures, with T decreasing from left to right. The inset shows non-optimal values of $\omega_t(\alpha)$ converging to the optimal curve. Each curve is for a specific inverse temperature $\beta = 0.3$ (light blue), $\beta = 0.4$ (brown) $\beta = 0.5$ (purple), $\beta = 0.6$ (red), $\beta = 0.7$ (green), $\beta = 1$ (yellow) and $\beta = 2$ (blue). (Graph created by Matteo Scandi).

system dynamics alone. This formed the the theoretical basis for the rest of the chapter, where I investigated the relation between work dissipation and fluctuations in the fully quantum setting.

In Section 4.3 I considered the effect of slow dynamics on the work statistics in general open quantum systems, analogous to the setting in which the classical work FDR holds. In this regime it can be shown that the system remains close to an instantaneous thermal state, with small deviations related to the slow but finite time driving of the Hamiltonian. Using these finite time corrections, I derived an expression for the quantum work fluctuations in the slow driving limit in Lemma 4.3.2, which decay to zero in the infinitely slow limit. It was found that the rate of work fluctuations is related to a type of quantum correlation function of the power operator. This result was then compared with an analogous expression previously obtained for the work dissipation,

which is also related to a form of quantum correlation function. However, the key insight was that in the quantum regime, the power operator describing the rate of change in energy will generally not commute with the instantaneous Hamiltonian of the system [154]. This in turn means that the correlation functions describing the respective work fluctuations and dissipation do not coincide in arbitrary open quantum systems. Bringing these observations together then resulted in the central result Theorem 4.3.1, which is the quantum generalisation of the work fluctuation-dissipation relation. The theorem demonstrates that work fluctuations are no longer proportional to dissipation due to this non-commutativity between energy and power. Instead, one obtains an additional quantum correction term. Intriguingly, in Theorem 4.3.2 it was found that this correction term is proportional to the average Wigner-Yanase-Dyson skew information encountered in Chapter 3, multiplied by an additional relaxation timescale determined by the open system dynamics. Due to the properties of the skew information, this term quantifies the excess non-classical fluctuations in power present along a generic protocol and leads to work fluctuations that are greater than the corresponding dissipation. I then demonstrated that the excess quantum power fluctuations necessarily lead to a non-Gaussian work distribution, in contrast to classical stochastic thermodynamics. Crucially this implies that the skew information can be witnessed directly from the work statistics.

This result, along with the quantum FDR, provides an unambiguous example of purely quantum effects present in thermodynamics below the nanoscale. Experimentally, one could detect these effects using ancilla-assisted measurement schemes that can be utilised to probe the work distribution via local measurements of the open system [144, 165, 166]. This should be contrasted with previous studies on the quantumness of $p(w)$ where weak measurements are needed to capture interference phenomena [167–170]. Notably, these different characterisations of quantum work all lead to the same fluctuation-dissipation relation due to the fact that there are no initial energy coherences in the initial state of the system within the framework. Instead, coherence effects emerge due to the intermediate dynamics of the open system rather than as a result of quantum measurement. In this sense we can interpret these results as a dynamical effect rather than measurement invasiveness. Since analytical expressions for the work distribution are difficult to obtain, it is significant that these result allow one to make predictions about the shape of $p(w)$ without having to actually compute its exact expression. It would be interesting to find an open system where $p(w)$ can be found exactly, and then to analyse to extent to which quantum power fluctuations change the shape of the distribution.

In Section 4.4, I investigated the geometric structure of the work statistics for

slow processes. While in classical stochastic thermodynamics the work dissipation and fluctuations are related to the classical Fisher-Rao metric [142], I demonstrated that this is no longer true in the quantum regime. Instead, the two quantities are actually related to different metrics over the Riemannian manifold of quantum thermal states, as shown in (4.97) and (4.99). Specifically, the dissipation is related to the so-called Kubo-Mori inner product [171] while the work fluctuations (4.99) are related to the symmetric covariance [172]. From a mathematical viewpoint the inequivalence of these metrics stems from the fact that within the manifold of quantum states there exist an infinite family of contractive metrics [173], which all reduce to the familiar Fisher-Rao metric in the classical commutative regime. Here we find that this mathematical structure has a direct physical consequence for quantum thermodynamics, as it leads to different notions of thermodynamic length and optimal paths. Much like the techniques used in classical geometric thermodynamics [140], one may solve the Euler-Lagrange equations to find optimal work extraction protocols for slow processes, with solutions corresponding to the relevant geodesics through the manifold of thermal states. However, since the metrics for dissipation and fluctuations do not generally coincide due to the non-commutativity between power and energy, the central realisation is that quantum processes cannot be simultaneously efficient with minimal fluctuations. This was made explicit in (4.107), where I showed that the presence of quantum power fluctuations leads to a non-vanishing trade-off between minimal dissipation and fluctuations. I then applied these results to the example of a driven harmonic oscillator, and derived expressions for the different optimal protocols and geodesics. This general framework should prove to be useful when designing quantum thermal machines, and the results of this chapter can be used to find the optimal compromise between both efficiency and reliability of any microscopic engine attempted to extract work.

This work opens the door to further directions of research. For example, throughout the chapter I have only considered open systems with a canonical thermal state as the fixed point of the dynamics. However, one expects that the techniques outlined here can be used to generalise the fluctuation-dissipation relation and thermodynamic geometry to more general non-equilibrium steady states. While this has been achieved in the classical setting [148], a fully quantum treatment has yet to be done. Since I have only considered the simplest case of single-parameter optimisation, it would be interesting to investigate how optimal thermodynamic paths depend on the interplay of multiple control parameters such as affinities, temperatures and Hamiltonian driving. Another interesting direction would be to investigate the connection between thermodynamic geometry and

quantum speed limits [88]. Recently a relation between thermodynamic length, entropy production and speed was established for classical Markovian systems [174]. One expects an analogous relationship for quantum systems, and it would be important to understand how quantum power fluctuations play a role in this context.

Bibliography

[1] A. Einstein, *Autobiographical Notes*, edited by P. A. Schilpp (Open Court Publishing, La Salle, IL, 1979).

[2] L. Landau and E. Lifshitz, *Course of Theoretical Physics: Statistical Physics*, Vol. 5 (Pergamon, Oxford, 1958) p. 76.

[3] A. B. Kolomeisky and M. E. Fisher, Ann. Rev. Phys. Chem. **58**, 675 (2007).

[4] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature **437**, 231 (2005).

[5] R. Kosloff and A. Levy, Ann. Rev. Phys. Chem. **65**, 365 (2013).

[6] P. Hänggi and P. Talkner, Nat. Phys. **11**, 108 (2015).

[7] C. Jarzynski, Phys. Rev. X **7**, 011008 (2017).

[8] K. Huang, *Lectures on Statistical Physics and Protein Folding* (World Scientific Publishing, Singapore, 2005).

[9] G. W. Ford, J. T. Lewis, and R. F. O'Connell, Phys. Rev. Lett. **55**, 2273 (1985).

[10] J. G. Kirkwood, J. Chem. Phys **3**, 300 (1935).

[11] H. J. Miller and J. Anders, Phys. Rev. E **95**, 062123 (2017).

[12] M. F. Gelin and M. Thoss, Phys. Rev. E **79**, 051121 (2009).

[13] Y. Subasi, C. H. Fleming, J. M. Taylor, and B. L. Hu, Phys. Rev. E **86**, 061132 (2012).

[14] E. P. Wigner and M. M. Yanase, J. Phys. Chem. **15**, 1084 (1963).

[15] H. J. D. Miller and J. Anders, Nat. Comm. **9**, 2203 (2018).

[16] P. Talkner, E. Lutz, and P. Hänggi, Phys. Rev. E **75**, 050102 (2007).

[17] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).

[18] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[19] G. E. Crooks, Phys. Rev. Lett. **99**, 100602 (2007).

[20] K. Sekimoto, Prog. Theor. Phys. Supp. **130**, 17 (1998).

[21] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[22] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).

[23] U. Seifert, Euro. Phys. J. E **34**, 26 (2011).

[24] U. Seifert, Rep. Prog. Phys **75**, 126001 (2012).

[25] R. Clausius, Annalen der Physik **201**, 353 (1865).

[26] G. Crooks, Phys. Rev. E **60**, 2721 (1999).

[27] T. Hatano and S. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).

[28] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).

[29] C. Jarzynski, Ann. Rev. Cond. M. Phys. **2**, 329 (2011).

[30] E. H. Feng and G. E. Crooks, Phys. Rev. Lett. **101**, 090602 (2008).

[31] J. M. R. Parrondo, C. V. den Broeck, and R. Kawai, N. J. Phys **11**, 073008 (2009).

[32] O. P. Saira, Y. Yoon, T. Tanttu, M. Möttönen, D. V. Averin, and J. P. Pekola, Phys. Rev. Lett. **109** (2012).

[33] A. I. Brown and D. A. Sivak, Phys. Rev. E **94** (2016).

[34] H. Kramers, Physica **7**, 284 (1940).

[35] P. Talkner, M. Campisi, and P. Hänggi, J. Stat. Mech. **2009**, P02025 (2009).

[36] S. Deffner and E. Lutz, Phys. Rev. Lett. **107**, 140404 (2011).

[37] P. Talkner and P. Hanggi, Phys. Rev. E **94**, 022143 (2016).

[38] U. Seifert, Phys. Rev. Lett. **116**, 020601 (2016).

[39] C. Jarzynski, J. Stat. Mech. **2004**, P09005 (2004).

[40] G. Ingold, P. Hänggi, and P. Talkner, Phys. Rev. E **79**, 061105 (2008).

[41] M. Esposito, K. Lindenberg, and C. Van den Broeck, N. J. Phys **12**, 013013 (2010).

[42] S. Hilt, S. Shabbir, J. Anders, and E. Lutz, Phys. Rev. E **83**, 25 (2011).

[43] M. Carrega, P. Solinas, M. Sassetti, and U. Weiss, Phys. Rev. Lett. **116** (2016).

[44] T. G. Philbin and J. Anders, J. Phys. A Math. Theo. **49**, 215303 (2016).

[45] P. Strasberg, G. Schaller, N. Lambert, and T. Brandes, N. J. Phys **18**, 073007 (2016).

[46] J. A. Thomas and T. M. Cover, *Elements of Information Theory* (Wiley, New York, 1991).

[47] H. Touchette, J. Stat. Phys. **159**, 987 (2015).

[48] M. J. Donald, J. Stat. Phys. **49**, 81 (1987).

[49] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[50] P. Strasberg and M. Esposito, Phys. Rev. E **95**, 062101 (2017).

[51] S. Vaikuntanathan and C. Jarzynski, EPL **87**, 60005 (2009).

[52] D. Reeb and M. M. Wolf, N. J. Phys **16**, 103011 (2014).

[53] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, EPL **82**, 50002 (2008).

[54] D. A. Hendrix and C. Jarzynski, J. Chem. Phys. **114**, 5974 (2001).

[55] T. Speck and U. Seifert, Phys. Rev. E **70**, 066112 (2004).

[56] P. Hänggi, G. L. Ingold, and P. Talkner, N. J. Phys **10**, 115008 (2008).

[57] N. Bohr, J. Chem. Soc. **1**, 349 (1932).

[58] B. Mandelbrot, IRE Trans . Inform. Theory **IT-2**, 190 (1956).

[59] G. D. J. Phillies, Am. J. Phys. **52**, 629 (1984).

[60] L. Velazquez and S. Curilef, Phys. Rev. A **31**, 3237 (1985).

[61] J. Lindhard, *The Lesson of Quantum Theory*, 1st ed., edited by J. de Boer, E. Dal and O. Ulfbeck (North-Holland, Amsterdam, 1986) p. 99.

[62] F. Schlögl, J. Phys. and Chem. Sol. **49**, 679 (1988).

[63] J. B. M. Uffink and J. van Lith, Found. Phys. **29**, 655 (1999).

[64] L. Velazquez and S. Curilef, Phys. Rev. A **31**, 2265 (2009).

[65] M. Falcioni, D. Villamaina, A. Vulpiani, A. Puglisi, and A. Sarracino, Am. J. Phys. **79**, 777 (2011).

[66] F. Giazotto, T. T. Heikkilä, A. Luukanen, A. M. Savin, and J. P. Pekola, Rev. Mod. Phys. **78**, 217 (2006).

[67] P. Zanardi, M. G. Paris, and L. Campos Venuti, Phys. Rev. A **78**, 042105 (2008).

[68] T. M. Stace, Phys. Rev. A **82**, 011611(R) (2010).

[69] A. De Pasquale, D. Rossini, R. Fazio, and V. Giovannetti, Nat. Comm. **7**, 12782 (2016).

[70] L. A. Correa, M. Mehboudi, G. Adesso, and A. Sanpera, Phys. Rev. Lett. **114**, 220405 (2015).

[71] M. Mehboudi, M. Moreno-Cardoner, G. De Chiara, and A. Sanpera, N. J. Phys **17**, 055020 (2015).

[72] M. Jarzyna and M. Zwierz, Phys. Rev. A **92**, 032112 (2015).

[73] S. Jevtic, D. Newman, T. Rudolph, and T. M. Stace, Phys. Rev. A **91**, 012331 (2015).

[74] F. Carlos, Luis Dias and Palacio, ed., *Thermometry at the Nanoscale* (The Royal Society of Chemistry, 2016).

[75] T. H. Johnson, F. Cosco, M. T. Mitchison, D. Jaksch, and S. R. Clark, Phys. Rev. A **93**, 053619 (2016).

[76] S. Campbell, M. Mehboudi, G. D. Chiara, and M. Paternostro, N. J. Phys **19**, 103003 (2017).

[77] P. P. Hofer, J. B. Brask, M. Perarnau-Llobet, and N. Brunner, Phys. Rev. Lett. **119**, 090603 (2017).

[78] A. Puglisi, A. Sarracino, and A. Vulpiani, Phys. Rep. **709**, 1 (2017).

[79] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Phot. **5**, 222 (2011).

[80] U. Marzolino and D. Braun, Phys. Rev. A **88**, 063609 (2013).

[81] L. A. Correa, M. Perarnau-Llobet, K. V. Hovhannisyan, S. Hernández-Santana, M. Mehboudi, and A. Sanpera, Phys. Rev. A **96**, 062103 (2017).

[82] M. Ludwig, K. Hammerer, and F. Marquardt, Phys. Rev. A **82**, 012333 (2010).

[83] S. Hilt, B. Thomas, and E. Lutz, Phys. Rev. E **84**, 031110 (2011).

[84] S. Luo, Phys. Rev. A **72**, 042110 (2005).

[85] X. Li, D. Li, H. Huang, X. Li, and L. C. Kwek, Euro. Phys J. D **64**, 147 (2011).

[86] I. Frérot and T. Roscilde, Phys. Rev. A **94**, 075121 (2016).

[87] I. Marvian and R. W. Spekkens, Nat. Comm. **5**, 3821 (2014).

[88] D. Paiva Pires, M. Cianciaruso, L. Céleri, G. Adesso, and D. Soares-Pinto, Phys. Rev. X **6**, 021031 (2016).

[89] M. Campisi, P. Talkner, and P. Hänggi, J. Phys. A Math. Theor. **42**, 392002 (2009).

[90] M. Campisi, D. Zueco, and P. Talkner, Chem. Phys. **375**, 187 (2010).

[91] S. Luo, Phys. Rev. A **73**, 022324 (2006).

[92] D. Girolami, Phys. Rev. Lett. **113**, 170401 (2014).

[93] I. Marvian, R. W. Spekkens, and P. Zanardi, Phys. Rev. A **93**, 052331 (2016).

[94] D. P. Pires, L. C. Céleri, and D. O. Soares-Pinto, Phys. Rev. A **91**, 042330 (2015).

[95] F. Hansen, Proc. Natl. Acad. Sci. USA **105**, 9909 (2008).

[96] M. Hayashi, *Quantum Information Theory: Mathematical Foundation* (Springer, Berlin, 2017).

[97] E. H. Lieb, Adv. Math. **11**, 267 (1973).

[98] S. Hilbert, P. Hänggi, and J. Dunkel, Phys. Rev. E **90**, 062116 (2014).

[99] M. Campisi, P. Talkner, and P. Hänggi, Phys. Rev. Lett. **102**, 210401 (2009).

[100] J. L. Lebowitz and L. Pastur, J. Phys. A **48**, 265201 (2015).

[101] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, Phys. Rev. X **4**, 031019 (2014).

[102] C. Gogolin and J. Eisert, Rep. Prog. Phys. **79**, 056001 (2016).

[103] A. Caldeira and A. Leggett, Physica A **121**, 587 (1983).

[104] H. Grabert, U. Weiss, and P. Talkner, Z. Phys. B. **55**, 87 (1984).

[105] B. L. Hu and J. P. Paz, Phys. Rev. D **45**, 2843 (1992).

[106] A. D. O. Connell, M. Hofheinz, M. Ansmann, R. C. Bialczak, M. Lenander, E. Lucero, M. Neeley, D. Sank, H. Wang, M. Weides, J. Wenner, J. M. Martinis, and A. N. Cleland, Nature **464**, 697 (2010).

[107] A. Lampo, S. H. Lim, M. Á. García-march, and M. Lewenstein, Quantum **1**, 30 (2017).

[108] T. G. Philbin, N. J. Phys **13**, 063026 (2011).

[109] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).

[110] P. Ullersma, Physica **32**, 27 (1966).

[111] C. W. Helstrom, J. Stat. Phys. **1**, 231 (1969).

[112] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[113] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[114] H. Nagaoka, *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (World Scientific, Singapore, 2005) pp. 125–132.

[115] A. Fujiwara, J. Phys. A **39**, 12489 (2006).

[116] S. Alipour, M. Mehboudi, and A. T. Rezakhani, Phys. Rev. Lett. **112**, 120405 (2013).

[117] U. Marzolino and T. Prosen, Phys. Rev. A. **90**, 062130 (2014).

[118] Armin Ulhmann, Rep. Math. Phys. **36**, 461 (1995).

[119] R. Bhatia and P. Rosenthal, Bull. London Math. Soc. **29**, 1 (1997).

[120] M. G. A. Paris, Int. J. Quant. Inf. **07**, 125 (2009).

[121] H. Araki, Ann. scient Ec. Norm. Supp **6**, 67 (1973).

[122] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications* (Springer, 2014).

[123] A. Rivas and S. Huelga, *Open Quantum Systems: An Introduction* (Springer Heidelberg, 2012).

[124] T. Mori and S. Miyashita, J. Phys. Soc. Japan **77**, 124005 (2008).

[125] L. Masanes and J. Oppenheim, Nat. Comm. **8**, 14538 (2017).

[126] H. Wilming and R. Gallego, Phys. Rev. X **7**, 041033 (2017).

[127] R. Uzdin, A. Levy, and R. Kosloff, Phys. Rev. X **5**, 031044 (2015).

[128] P. Kammerlander and J. Anders, Sci. Rep. **6**, 22174 (2016).

[129] S. Luo, Phys. Rev. Lett. **91**, 180403 (2003).

[130] P. Neumann, I. Jakobi, F. Dolde, C. Burk, R. Reuter, G. Waldherr, J. Honert, T. Wolf, A. Brunner, J. Wrachtrup, J. H. Shim, D. Suter, H. Sumiya, and J. Isoya, Nano Lett. **13**, 2738 (2013).

[131] G. Kucsko, P. C. Maurer, N. Y. Yao, M. Kubo, H. J. Noh, P. K. Lo, and H. Park, Nature **500**, 54 (2013).

[132] J. O. González, L. A. Correa, G. Nocerino, J. P. Palao, D. Alonso, and G. Adesso, Open Syst. Inf. Dyn. **24**, 1740010 (2017).

[133] P. P. Hofer, M. Perarnau-Llobet, L. D. M. Miranda, G. Haack, R. Silva, J. B. Brask, and N. Brunner, N. J. Phys **19**, 123037 (2017).

[134] J. Iles-Smith, N. Lambert, and A. Nazir, Phys. Rev. A **90**, 032114 (2014).

[135] D. Newman, F. Mintert, and A. Nazir, Phys. Rev. E **95**, 032139 (2017).

[136] K. V. Hovhannisyan and L. A. Correa, Phys. Rev. B **98**, 045101 (2018).

[137] J. Hermans, J. Phys. Chem. **95**, 9029 (1991).

[138] R. H. Wood and W. C. F. Muhlbauer, J. Phys. Chem. **95**, 6670 (1991).

[139] S. I. Amari and H. Nagaoka, *Methods of information geometry* (American Mathematical Soc., 2007).

[140] G. Ruppeiner, Phys. Rev. A **20**, 1608 (1979).

[141] P. Salamon and R. S. Berry, Phys. Rev. Lett. **51**, 1127 (1983).

[142] D. A. Sivak and G. E. Crooks, Phys. Rev. L **108**, 190602 (2012) (2012).

[143] M. Esposito, U. Harbola, and S. Mukamel, Rev. Mod. Phys. **81**, 1665 (2009).

[144] R. Dorner, S. R. Clark, L. Heaney, R. Fazio, J. Goold, and V. Vedral, Phys. Rev. Lett. **110**, 230601 (2013).

[145] L. Mazzola, G. De Chiara, and M. Paternostro, Phys. Rev. Lett. **110**, 230602 (2013).

[146] T. B. Batalhão, A. M. Souza, L. Mazzola, R. Auccaise, R. S. Sarthour, I. S. Oliveira, J. Goold, G. De Chiara, M. Paternostro, and R. M. Serra, Phys. Rev. Lett. **113**, 140601 (2014).

[147] M. Scandi and M. Perarnau-Llobet, arXiv:1810.05583 (2018).

[148] D. Mandal and C. Jarzynski, J. Stat. Mech. **2016**, 063204 (2016).

[149] A. P. Solon and J. M. Horowitz, Phys. Rev. Lett. **120**, 180605 (2018).

[150] P. Talkner and P. Hänggi, J. Phys. A **40**, F569 (2007).

[151] P. Talkner, P. Hänggi, and M. Morillo, Phys. Rev. E **77**, 051131 (2008).

[152] M. Campisi, P. Talkner, and P. Hänggi, Phys. Rev. Lett. **105**, 10 (2010).

[153] S. An, J.-N. Zhang, M. Um, D. Lv, Y. Lu, J. Zhang, Z.-Q. Yin, H. T. Quan, and K. Kim, Nat. Phys. **11**, 193 (2014).

[154] S. Suomela, P. Solinas, J. P. Pekola, J. Ankerhold, and T. Ala-Nissila, Phys. Rev. B **90**, 094304 (2014).

[155] C. Jarzynski, H. T. Quan, and S. Rahav, Phys. Rev. X **5**, 031038 (2015).

[156] B. Prasanna Venkatesh, G. Watanabe, and P. Talkner, N. J. Phys **17**, 075018 (2015).

[157] H. Carmichael, *An Open System Approach To Quantum Optics* (Springer, 1991).

[158] H.-P. Breuer, F. Petruccione, *et al.*, *The theory of open quantum systems* (Oxford University Press on Demand, 2002).

[159] T. Albash, S. Boixo, and D. A. Lidar, N. J. Phys **14**, 123016 (2012).

[160] T. L. Boullion and P. L. Odell, *Generalised inverse matrices* (Wiley-Interscience, New York, 1971).

[161] V. Cavina, A. Mari, and V. Giovannetti, Phys. Rev. Lett. **119**, 050601 (2017).

[162] F. Kubo and T. Ando, Math. Ann. **224**, 205 (1980).

[163] J. Briët and P. Harremoës, Phys. Rev. A **79**, 1 (2009).

[164] H. Spohn, J. Math. Phys. **19**, 1227 (1978).

[165] A. J. Roncaglia, F. Cerisola, and J. P. Paz, Phys. Rev. Lett. **113**, 250601 (2014).

[166] P. Solinas, H. J. D. Miller, and J. Anders, Phys. Rev. A **96**, 052115 (2017).

[167] A. E. Allahverdyan, Phys. Rev. E **90**, 032137 (2014).

[168] P. Solinas and S. Gasparinetti, Phys. Rev. E **92**, 042150 (2015).

[169] H. J. D. Miller and J. Anders, N. J. Phys **19**, 062001 (2017).

[170] R. Blattmann and K. Mølmer, Phys. Rev. A **96**, 012115 (2017).

[171] D. Petz, J. Math. Phys. **35**, 780 (1994).

[172] P. Gibilisco and T. Isola, Ann. Inst. Stat. Mech. **59**, 147 (2007).

[173] D. Petz, Lin. Alg. App. **244**, 81 (1996).

[174] S. Ito, Phys. Rev. Lett. **121**, 030605 (2018).